METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

Vol. VIII - N. 1 - 2 **30 - VI - 1929**

SOMMARIO — SOMMAIRE — CONTENTS — INHALT

ROMA
AMMINISTRAZIONE DEL « METRON »
R. UNIVERSITÀ — ISTITUTO DI STATISTICA
:: :: E POLITICA ECONOMICA :: ::

ARTICOLI GIUNTI ALLA RIVISTA CHE VERRANNO PUBBLICATI NEI PROSSIMI NUMERI.
(*Secondo l'ordine d'arrivo*)

ARTIKEL DIE AN DIE ZEITSCHRIFT ANGELANGT SIND UND WELCHE IN DEN NACHFOLGENDEN NUMMERN ERSCHEINEN WERDEN.
(*Nach der Reihenfolge des Eingangs*)

ARTICLES REÇUS PAR LA REVUE ET À PARAÎTRE PROCHAINEMENT.
(*D'après la date de reception*)

ARTICLES RECEIVED BY THE REVIEW WHICH WILL BE PUBLISHED IN FUTURES ISSUES.
(*According to date of recept*)

V. A. Nekrassoff. *Nomography in applications of Statistics.*

C. Gini. *Sul massimo degli indici di variabilità assoluta e sulle sue applicazioni agli indici di variabilità relativa e al rapporto di concentrazione.*

R. Roy. *La demande dans ses rapports avec la repartition des revenus.*

J. C. Irwin, M. A., M. Sc. *On the frequency distribution of the means of samples from populations of certain of Pearson's types.*

C. GINI e L. GALVANI

DI TALUNE ESTENSIONI DEI CONCETTI DI MEDIA AI CARATTERI QUALITATIVI

SOMMARIO

I. — INTRODUZIONE

Corrispondenze — Seriazioni e serie statistiche

1. — Quando in un insieme di fenomeni si considerano simultaneamente due dei loro caratteri X ed Y, e si tiene conto delle modalità assunte dal carattere Y corrispondentemente a quelle assunte da X, si ottiene un « sistema di elementi corrispondenti » o, più semplicemente, una « corrispondenza ». I due caratteri possono essere entrambi quantitativi, e entrambi qualitativi, o infine uno quantitativo e l'altro qualitativo. Così se un insieme di redditieri si distribuisce in classi a seconda dell'altezza del reddito (carattere X) e in ciascuna classe si determina il numero dei redditieri, o l'ammontare globale dei redditi, o il reddito medio, ecc. (carattere Y) si ha una corrispondenza fra due caratteri quantitativi. Se una certa popolazione si distribuisce in classi a seconda della religione, della razza o della lingua parlata (carattere X) e in ciascuna classe si tiene conto del colore predominante degli occhi o della forma o del colore dei capelli (carattere Y), si ottiene una corrispondenza fra due caratteri qualitativi. Se i matrimoni avvenuti in un certo intervallo di tempo si classificano a seconda del giorno della settimana in cui furono celebrati (carattere X) e per ciascuna classe si considera il numero dei matrimoni, o l'età media degli sposi, o l'ammontare complessivo dei beni dotali delle spose (carattere Y), la corrispondenza ha luogo fra un carattere (X) qualitativo ed un carattere (Y) quantitativo. Finalmente se un gruppo di reclute si classifica a seconda dell'altezza, o del peso o del perimetro toracico (carattere X), e si determina la regione (carattere Y) da cui proviene prevalentemente ciascuna classe, si ha una corrispondenza fra un carattere (X) quantitativo, ed uno (Y) qualitativo.

In tutti questi casi i caratteri X ed Y sono, in un certo modo, paragonabili a quelle che nell'analisi si considerano rispettivamente come variabile indipendente e variabile dipendente o funzione, in quanto il carattere X, pur non essendo di solito variabile arbitrariamente, è quello rispetto al quale si effettua la distribuzione dello insieme considerato in classi, mentre il carattere Y varia dipenden-

temente dalle classi così formate. Ma è ovvio che il posto concetto di corrispondenza non va confuso con quello di dipendenza funzionale di Y da X, in senso analitico, per la ragione che X ed Y possono essere delle qualità e non delle quantità.

Per indicare che in una corrispondenza si considerano i valori o le modalità che il carattere Y può assumere in dipendenza dei valori o modalità di X si potrà scrivere:

$$Y = S((X))$$

Se Y è un carattere *quantitativo statistico* la corrispondenza $Y = S((X))$ costituisce ciò che si dice una «serie statistica».

Fra queste si dicono più particolarmente *seriazioni* quelle nelle quali il carattere X è quantitativo, mentre si continuano senz'altro a chiamare *serie* quelle in cui il carattere X è qualitativo. Il carattere X si dice «variabile» se quantitativo e «mutabile» se qualitativo.

Infine, se si assume come carattere Y il numero dei termini di ciascuna classe ottenuta distribuendo la totalità dei fenomeni considerati a seconda delle modalità, quantitative o qualitative, di X, si hanno rispettivamente le seriazioni e le serie «di frequenze».

2. — Le corrispondenze più spesso prese in esame sono quelle in cui Y sia quantitativo; ma anche le altre, per le quali il carattere Y sia qualitativo, possono interessare lo statistico. Se la considerazione di queste ultime è meno frequente, ciò dipende soprattutto dalla impossibilità o dalla difficoltà di applicare ad esse il concetto di misura, e conseguentemente tutti quegli strumenti di calcolo i quali, nei campi in cui ne sia lecito l'uso, consentono una più profonda ed esauriente investigazione. Se il carattere X è quantitativo ed Y qualitativo, la difficoltà può talora essere superata, o quanto meno ridotta, considerando X come dipendente (inversamente) da Y. Per esempio, invece di classificare delle reclute a seconda dell'altezza e di determinare poi la regione da cui proviene prevalentemente ciascuna classe, può essere lo stesso, agli effetti della ricerca di una eventuale relazione fra regione e altezza, eseguire la classificazione di quelle reclute per regioni e calcolare poi l'altezza media (o la differenza media, o un altro indice di variabilità delle altezze) per le reclute di ciascuna regione.

Così anche, se tanto X che Y fossero caratteri qualitativi, una eventuale relazione fra X ed Y potrebbe essere messa in luce attraverso un carattere Z quantitativo, di cui X ed Y si potessero riguardare come funzioni. Ma, in generale, quelle corrispondenze in cui

il carattere Y che si considera come funzione sia qualitativo, non possono assoggettarsi a profonda investigazione. Ci basti averne qui fatto cenno per completare il quadro delle corrispondenze; e volgiamoci a considerare le sole seriazioni e serie statistiche, cioè le sole corrispondenze in cui Y sia un carattere quantitativo statistico.

CONCETTI FONDAMENTALI PER LO STUDIO DELLE SERIE STATISTICHE

3. — Per la stessa ragione or ora accennata, la teoria delle seriazioni, come quella che si è avvantaggiata della possibilità di applicare il concetto di misura a ciascuno dei due caratteri quantitativi che in esse intervengono, ha compiuto progressi più rapidi che non la teoria delle serie, nelle quali soltanto il carattere Y (funzione) è immediatamente assoggettabile a misura.

È vero che il GALTON (1), il PEARSON (2), lo YULE (3), il BENINI (4) ed altri si occuparono, da diversi punti di vista, di estendere alle serie statistiche alcuni dei concetti e dei procedimenti proprî delle seriazioni; ma soltanto di recente (5) la trattazione sistematica delle serie è stata affrontata in tutta la sua ampiezza e generalità ed ha compiuto progressi decisivi coordinandosi ai seguenti criterî:

1°) Distinzione razionale delle serie in base ai diversi tipi di caratteri qualitativi da cui provengono.

2°) Applicazione del concetto di misura non già alle modalità – ciò che sarebbe quasi sempre impossibile – ma alle «diversità»

(1) *Statistics by Intercomparison with Remarks on the Law of Frequency of Error* « The London Edinburgh and Dublin Philosophical Magazine and Journal of Science », Vol. XIL, IV series, 1875.

(2) K. PEARSON e A. LEE. *On the inheritance of caracters not capable of exact quantitative measurement* « Phil. Trans. of the Roy. Soc. of London », Serie A, CXCV, 1900; K. PEARSON. *On the correlation of caracters not quantitatively measurable.* Ibid.

(3) *On the association of attributes in statistics* « Phil. Trans. Roy. Soc. », 1900; *On a property which holds good for all grouping... with applications to the study of contingency tables for the inheritance of unmeasured qualities* « Proc. Roy. Soc. », 1906.

(4) *Principi di demografia*, Firenze, 1901, pag. 129, 130.

(5) C. GINI. *Variabilità e mutabilità* in « Studi economico-giuridici della Regia Università di Cagliari », 1912; *Di una estensione del concetto di scostamento medio e di alcune applicazioni alla misura della variabilità dei caratteri qualitativi*, « Atti del R. Istituto Ven. di S. L. A. », 1917-18, Tomo LXXVII, Parte II.

che presentano fra di loro le varie modalità del carattere qualitativo che dà origine a ciascuna serie, il che riesce assai più facile.

La distinzione di cui al primo punto (1), necessaria per dare alla teoria basi solide e precise, riposa sulla osservazione che vi sono dei caratteri qualitativi le cui modalità presentano un ordine naturale di successione (gradi di una gerarchia, giorni della settimana, direzioni del vento, ecc.) e ve ne sono altri le cui modalità non presentano ordine alcuno (razza, religione, professione esercitata, ecc.). Le serie statistiche corrispondenti a queste due diverse classi di caratteri sono state dette rispettivamente *ordinate* e *non ordinate*. A loro volta le serie ordinate si sono distinte in *rettilinee* e *cicliche*, secondo che il carattere da cui dipendono sia *rettilineo*, cioè ammetta naturalmente due modalità estreme (grado di una gerarchia, colore degli occhi dal celeste al bruno), oppure sia *ciclico*, cioè non ammetta due modalità estreme o tutt'al più le ammetta soltanto in base ad una convenzione (p. es. giorno della settimana, direzione sulla rosa dei venti). Infine, di serie non ordinate sono possibili diverse varietà, per la circostanza che (non tutte nel loro complesso, ma) alcune delle modalità del carattere possono ben presentare un certo ordine di successione (p. es. in una classificazione per religioni, le varie chiese cristiane protestanti); ma, come caso tipico, si considera quello estremo, cioè quello in cui due qualunque delle modalità del carattere si debbono reputare avere sempre fra di loro una stessa diversità: serie di questo tipo si sono dette *sconnesse*. Così la serie ottenuta col distribuire la popolazione mondiale a seconda della lingua parlata è sconnessa; precisamente una serie sconnessa di frequenze.

Quanto al secondo concetto posto a fondamento della trattazione sistematica della serie, consistente nell'esprimere numericamente le diversità fra le modalità di un carattere qualitativo, esso, insieme col precedente, fu immediatamente fecondo di risultati. In un primo tempo fu possibile (2) definire ed esprimere in formule: per le serie rettilinee tutti gli indici di mutabilità (corrispondenti

(1) C. GINI. *Variabilità*, ecc. già cit., pag. 114.

(2) C. GINI. *Variabilità*, ecc. già cit. Sull'idea di far corrispondere alle qualità di un carattere i termini di una progressione aritmetica allo scopo di paragonare certe proprietà delle serie e in particolare per giungere ad una determinazione della modalità media delle serie (e non già per misurarne la mutabilità), V. nota ivi, pag. 121. È in essa ricordato come il BEDDOE, per esempio, calcolasse l'indice di nigrescenza, facendo corrispondere al colore bruno dei capelli il numero 2, al castagno l'1, al rosso e al biondo l'1.

agli indici di variabilità delle seriazioni) e cioè la differenza media semplice e con ripetizione e gli scostamenti semplici e quadratici dalla media aritmetica e dalla mediana; per le serie cicliche la differenza media semplice e con ripetizione; per le sconnesse gli scostamenti medî semplice e quadratico, la differenza media semplice e con ripetizione. Inoltre venne in tutti questi casi studiato l'effetto esercitato sui valori degli indici di mutabilità dal raggruppare in classi le modalità qualitative e furono stabiliti i criteri per il confronto degli indici di mutabilità di più serie. Successivamente, (1) estendendo i concetti di scostamento medio semplice e quadratico dalla media, senza e con ripetizione, in modo da renderli indipendenti dalla conoscenza della media aritmetica (della mediana o di altro valore medio della serie), fu possibile definire tali scostamenti anche per le serie cicliche e sconnesse, per le quali appunto non si sono finora definite nè la media nè la mediana: e fu dimostrata la legittimità e la opportunità di tali estensioni, provando che nel caso delle seriazioni esse vengono a coincidere con i consueti scostamenti dalla media aritmetica. Vennero anche determinate le relazioni che intercedono fra i valori degli scostamenti medî e quelli delle differenze medie, e venne data ragione della eventualità di ottenere risultati diversi secondo che si impieghino i primi o le seconde in serie di diverso o anche del medesimo tipo.

Nel frattempo progrediva anche la teoria delle relazioni fra i caratteri sia quantitativi che qualitativi (2), mercè la distinzione dei concetti di connessione e di concordanza, confusi solitamente sotto il termine comune di correlazione e mercè la creazione di indici semplici e quadratici convenienti per la misura di ciascuno di essi. Finalmente il PIETRA (3), fermandosi particolarmente allo studio delle serie cicliche, e determinando in primo luogo il minimo ed il massimo della somma dei valori assoluti delle differenze (e dei quadrati delle differenze) fra le modalità corrispondenti di due serie cicliche,

(1) C. GINI. *Di una estensione del concetto di scostamento medio*, ecc. già cit.

(2) C. GINI. *Di una misura della disuguaglianza fra due gruppi di quantità e sua applicazione allo studio delle relazioni statistiche*, « Atti del R. Istituto Ven. S. L. A. », 1914-15, Tomo LXXIV, Parte II; *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione*, Ibid. *Nuovi contributi alla teoria delle relazioni statistiche*, Ibid.

(3) *The theory of statistical relations with special reference to cyclical series*, « Metron », Volume IV, 1925.

riuscì a definire gli indici semplice e quadratico di dissomiglianza fra due di tali serie ed a completare per le serie stesse le teorie della connessione e della concordanza.

Condizioni sufficienti alla ricerca della media aritmetica e degli indici di variabilità

4. — Data così una succinta idea della fecondità dei due principî posti a base dello studio delle serie, vale la pena di soffermarsi a considerare l'essenza e la portata del secondo di essi, consistente dunque nel *misurare* non le varie modalità di un carattere qualitativo, ma le *diversità* fra di loro intercedenti.

Supponiamo di avere a che fare con un carattere quantitativo, p. es. «lunghezza rispetto ad una unità fissata» di una classe di segmenti che sopra una retta hanno una certa origine e un certo verso. Siano a, b, c,... le loro misure, e si consideri la trasformazione lineare:

$$f\ (x) = px + q,$$

dalla quale risulterà

$$f\,(a) = a',\ f\,(b) = b',\ f\,(c) = c', \ldots$$

Mentre la corrispondenza fra i varî segmenti ed i numeri a, b, c,... era biunivoca e proporzionale, la corrispondenza fra i segmenti stessi ed i numeri a', b', c',... è ancora biunivoca, ma non proporzionale (a meno che non sia $q = o$); perciò non si può dire, nel senso ordinario della parola, che a', b', c',... siano «misure» di quei certi segmenti. Peraltro, dal momento che il vantaggio principale che si consegue sostituendo alle grandezze di una classe i numeri che le misurano è quello di fare più semplicemente sui numeri certe operazioni in luogo di altre che si dovrebbero fare sulle grandezze per determinarne una incognita, potrà ben darsi che per alcune classi di operazioni, sia indifferente operare sui numeri a, b, c,... o sui numeri a', b', c',... agli effetti di determinare quella certa incognita. In particolare se φ è una operazione commutabile con f, cioè se:

$$\varphi\ (a', b', c') = \varphi \{ f\,(a), f\,(b), f\,(c), \ldots \} = f \{ \varphi\ (a, b, c, \ldots) \}$$

le espressioni $\varphi\,(a', b', c', \ldots)$ e $\varphi\,(a, b, c, \ldots)$ saranno numericamente diverse ma corrisponderanno ad una stessa grandezza. L'operazione di media aritmetica soddisfa appunto alla condizione di essere commu-

tabile con la f perchè essendo A ed A' le medie rispettive degli n numeri $a, b, c, \ldots$ e degli n numeri $a', b', c', \ldots$ si ha:

$$A' = \frac{f(a) + f(b) + f(c) + \ldots}{n} = \frac{(pa + q) + (pb + q) + \ldots}{n} =$$

$$= p\,\frac{a + b + c + \ldots}{n} + q = f\left(\frac{a + b + c + \ldots}{n}\right) = f(A)$$

e pertanto *agli effetti della ricerca della media aritmetica di segmenti o più in generale di grandezze aventi per misure* a, b, c,... *è indifferente operare su questi numeri o sui loro trasformati lineari mediante la* f (x) = px + q (1).

Di qui, osservando che $a' - A' = p\,(a - A)$, ecc., deriva subito che se gli usuali indici di variabilità del sistema di numeri $a, b, c, \ldots$ sono: Δ (differenza media), Δ_R (differenza media con ripetizione), η_A (scostamento semplice dalla media) e ${}^2\eta_A$ (scostamento quadratico medio dalla media), mentre Δ', Δ'_R, $\eta'_{A'}$, ${}^2\eta'_{A'}$ sono i corrispondenti indici per il sistema $a', b', c', \ldots$ si ha:

$$\Delta' = p\,\Delta \qquad\qquad \Delta'_R = p\,\Delta_R$$

$$\eta'_{A'} = p\,\eta_A;\quad {}^2\eta'_{A'} = \sqrt{\frac{(a' - A')^2 + (b' - A')^2 + \ldots}{n}} =$$

$$= \sqrt{\frac{p^2\,(a - A)^2 + p^2\,(b - A)^2 + \ldots}{n}} = p \times {}^2\eta_A$$

Dunque, *se sui numeri* a, b, c,. .. *si effettua una trasformazione lineare, i loro usuali indici di variabilità risultano moltiplicati per il coefficiente della variabile.*

In particolare *la traslazione*

$$f\,(x) = x + q$$

lascia invariato (la grandezza corrispondente alla media aritmetica e) *ciascuno degli indici di variabilità* Δ, Δ_R, η_A, ${}^2\eta_A$.

Ora, l'eseguire una trasformazione lineare sulle misure di una classe di grandezze, è lo stesso che eseguire un cambiamento della

(1) Cfr. L. March. *L'analyse de la variabilité,* « Metron » 1, 6, 1926. Non si potrebbe dire altrettanto per le medie geometrica ed armonica, a meno che non sia $q = O$. Invece è appena il caso di avvertire che la mediana e la moda di più modalità quantitative non cambiano per effetto di una trasformazione lineare, perchè dipendono dall'ordine di successione e dalle frequenze di tali modalità e non dalla loro grandezza.

origine e dell'unità di misura; ossia è lo stesso che fare corrispondere ad una delle grandezze, p. es. G, un numero qualunque μ, misurare la differenza fra ogni altra grandezza H e quella prima G mediante una unità arbitraria, e infine far corrispondere ad H il numero

$$\nu = \mu + \text{misura di } (H-G) \quad (1)$$

Con ciò a coppie di grandezze equidifferenti corrisponderanno coppie di numeri equidifferenti.

Volendo dunque determinare la media aritmetica e gli usuali indici di variabilità di un sistema di modalità di un carattere quantitativo, non è necessario conoscere le misure delle varie modalità, ma è sufficiente stabilire una corrispondenza biunivoca fra tali modalità ed i numeri reali, con la sola condizione che a coppie di modalità equidifferenti, corrispondano coppie di numeri equidifferenti (2).

È questo appunto il principio che è stato utilmente sfruttato dal GINI per lo studio delle serie statistiche e che ha condotto la loro teoria quasi allo stesso grado di sviluppo di quella riguardante le seriazioni.

Posto, infatti, che di un carattere qualitativo X (considerato in una serie statistica) non sia possibile misurare le varie modalità, ma si sappiano invece in qualche modo *misurare* le *diversità* fra le modalità stesse, ne risulterà la possibilità di far corrispondere a tali modalità dei numeri, con la condizione che a coppie di modalità *equidiverse* corrispondano coppie di numeri *equidifferenti*. Naturalmente non basta questo per concludere senz'altro che alle serie sia applicabile il calcolo come alle seriazioni; ma poichè per la ricerca

(1) Per esempio, rispetto ad una certa unità le misure dei segmenti OA ed OB siano 5 ed 8. Facciamo corrispondere al segmento OA un numero qualunque, per es.: 11, misuriamo la differenza OB-OA con una unità arbitraria, e sia 4 la misura ottenuta, cosicchè al segmento OB verrà a corrispondere il numero $11 + 4 = 15$. La trasformazione cercata, della forma $y = px + q$, risulta definita dalle equazioni $5p + q = 11$, $8p + q = 15$, che forniscono $p = 4/3$, $q = 13/3$, onde $y = (4/3)x + (13/3)$.

(2) Le misure di alcune modalità di un carattere X quantitativo siano 5, 8, 15, 25, 32. La media aritmetica è $A = 17$; gli usuali indici di variabilità sono: $\Delta = 7{,}1$; $\Delta_R = 5{,}68$; $\eta_A = 9{,}2$ $_2\eta_A = 10{,}18$. Eseguendo la trasformazione, di cui alla nota precedente, $y = (4/3)x + (13/3)$, ai valori 5, 8, 15, 25, 32 corrispondono i valori 11, 15, 24 1/3, 37 2/3, 47. Di questi la media aritmetica è $A' = 27 = (4/3)\,17 + (13/3)$, cioè è la trasformata, mediante la stessa trasformazione, della media A. Inoltre risulta $\Delta' = 9{,}466 = 4/3 \cdot 7{,}1$; $\Delta'_R = 7{,}5733 = 4/3 \cdot 5{,}68$; $\eta'_{A'} = 12{,}266... = 4/3 \cdot 9{,}2$; $_2\eta'_{A'} = 13{,}57 = 4/3 \cdot 10{,}18$.

della media aritmetica A e degli accennati indici di variabilità Δ, Δ_R, ${}^1\eta_A$, ${}^2\eta_A$, basta, nel caso delle seriazioni, la misura delle differenze fra le modalità del carattere avente il ruolo di variabile indipendente, così abbiamo già una solida base per applicare il calcolo anche alle serie statistiche, salvo poi a determinare le particolarità di tale applicazione nel caso delle serie rettilinee, delle cicliche e delle sconnesse.

In base alla misura di quelle diversità è altresì possibile determinare, come si vedrà, la mediana M di una serie e conseguentemente gli scostamenti semplice e quadratico ${}^1\eta_M$, ${}^2\eta_M$ da tale mediana.

Insomma, l'adattabilità del calcolo alle serie è meno diretta che per le seriazioni, ma non è impossibile.

Considereremo anzitutto le serie ordinate.

Come si possono, dunque, « misurare » le diversità tra le modalità di un carattere X qualitativo ordinato, e quindi porre in corrispondenza tali modalità con dei numeri ?

Come si può far uso di tali numeri per definire e determinare le modalità medie e gli indici di mutabilità di una serie data in funzione di quel carattere X ?

Lo vedremo separatamente per le serie rettilinee e per le serie cicliche.

Precisazione del concetto di « modalità » di un carattere qualitativo ordinato

5. — Ma ci conviene, anzitutto, meglio precisare il concetto di « modalità » di un carattere qualitativo ordinato.

Vi sono dei caratteri qualitativi ordinati i quali risultano di per sè stessi distinti in varie successive modalità: tali sono, p. es., i diversi gradi di una certa gerarchia (carattere rettilineo) e i valori dell'ultima cifra nella successione dei numeri naturali (carattere ciclico). Così nel nostro ordinamento militare al grado di capitano segue quello di maggiore e fra i due gradi vi è soluzione di continuità, cioè fra i due non è compreso nessun grado intermedio. Così pure, se nella successione dei numeri naturali consideriamo come un carattere il valore o, ciò che è lo stesso, la forma dell'ultima cifra, la quale assume in ordine ciclico le varie modalità 0, 1, . . . 9, 0, 1, 2, . . ., ecco che tali modalità risulteranno naturalmente distinte fra di loro,

in quanto fra le due classi dei numeri naturali terminanti rispettivamente, poniamo, con le cifre 3 e 4 non esiste nessun'altra classe intermedia. Naturalmente in ciascuno di tali caratteri anche il numero delle modalità è perfettamente determinato.

Ma vi sono altri caratteri qualitativi ordinati, e sono anzi la maggior parte, nei quali le modalità successive sono naturalmente infinite e possono essere raggruppate in un numero finito di modalità opportunamente nominate o, come si potrebbe dire, modalità – classi, soltanto in seguito ad una convenzione esplicita o sottintesa per la quale ciascuna di queste comprenda una classe infinita delle prime; al variare della convenzione potranno variare, sia le modalità nominate, sia il loro numero.

Per esempio, del carattere rettilineo « colore dei capelli » si potrà convenire di considerare le successive quattro modalità: biondo, rosso, castano, nero, ciascuna delle quali avrà una certa estensione, in modo da comprendere tutte le infinite sfumature che, per comune consenso, rientrano nella stessa denominazione di biondo, rosso, ecc. Ma nulla vieta di stabilire una diversa convenzione per la quale le successive modalità dello stesso carattere siano, p. es. denominate: biondo, biondo-rosso, rosso, rosso-castano, castano, castano-bruno, nero; ciascuna di tali modalità avrà questa volta una estensione minore di prima, *anche quando la denominazione sia la stessa*. Similmente, il carattere ciclico « direzione del vento » in un dato punto della superficie terrestre può essere distinto in quattro sole modalità denominate Nord, Est, Sud, Ovest, ciascuna delle quali comprenderà tutte le direzioni contenute in un certo quadrante. Ma lo stesso carattere si potrà benissimo distinguere in 8 modalità denominate nel modo consueto *N*, *N-E*, *E*, *S-E*, *S*, *S-W*, *W*, *N-W*, ciascuna delle quali comprenderà tutte le direzioni contenute in un certo settore (ottante); anche qui è variata l'estensione di ciascuna modalità, cosicchè la modalità *N* che aveva prima una estensione di 90^o, ha ora un'estensione di 45^o. Caratteri cosiffatti si potranno dire « continui », mentre i primi, per contrapposto, si diranno « discreti ».

Dato che la distinzione dei caratteri continui in modalità-classi è convenzionale, potrà talora convenire di sostituire, anche nel corso di una stessa indagine, ad un certo sistema di modalità un nuovo sistema, generalmente più numeroso del primo; e se ne vedranno nel seguito diversi esempi. Non è poi detto che tutte le modalità di un carattere continuo abbiano la stessa estensione.

In tutti i casi di considerazione di un certo carattere continuo supporremo di averlo preventivamente distinto in un numero finito di modalità-classi opportunamente nominate, e queste diremo senza altro « modalità » di quel carattere.

II. — SERIE RETTILINEE

Principi per l'applicazione dei procedimenti di calcolo alle serie rettilinee

6. — Benchè la sistematica applicazione del calcolo alle serie rettilinee sia già stata compiutamente eseguita (Cfr. note al § 3), gioverà nondimeno esporne con maggiore ampiezza i principî per vedere se e fino a qual punto gli stessi principî possano impiegarsi per le serie cicliche e per le sconnesse.

Sia X un carattere rettilineo, le cui modalità si seguiranno in due sensi od ordini naturali contrarî; diremo, di tali sensi, uno positivo e l'altro negativo. Di due modalità A e B si potranno quindi considerare due diversità: quella da A a B e quella da B ad A, che indicheremo rispettivamente con (AB) e con (BA); fra le due diversità si dirà positiva quella in cui le modalità si seguono nell'ordine già fissato come positivo, e negativa l'altra. Diremo pure che tali diversità sono opposte fra di loro, e scriveremo:

$$(AB) = -(BA) \quad , \quad (BA) = -(AB)$$

Naturalmente, qualsiasi ambiguità circa la diversità di A e B cessa con lo stabilire quale fra le due modalità si debba considerare come prima; mancando tale designazione si potrà parlare soltanto di diversità in valore assoluto fra A e B, che si indicherà indifferentemente con $|AB|$ o con $|BA|$.

Essendo A, B, C tre modalità *qualunque* di X, diremo che la diversità da A a C è la somma di quella da A a B e da B a C, e scriveremo:

$$(AC) = (AB) + (BC)$$

e diremo altresì che la diversità da A a B è la differenza fra quelle da A a C e da B a C, scrivendo:

$$(AB) = (AC) - (BC)$$

da cui

$$(AB) = (CB) - (CA)$$

Dunque la diversità da una prima ad una seconda modalità è la differenza fra la diversità da una terza modalità alla seconda e la diversità da questa terza alla prima.

Se $(AC) = (AB) + (BC)$ ed è (BC) positiva, diremo che la diversità da A a C è maggiore di quella da A a B, ossia $(AC) > (AB)$.

La diversità dalla prima all'ultima modalità di X in un certo ordine si può dire *ampiezza dell'intervallo di mutabilità* di X. Tale ampiezza è la somma delle diversità dalla 1ª alla 2ª, dalla 2ª alla 3ª e così fino all'ultima modalità.

È appena necessario osservare che tutte queste convenzioni sono suggerite dall'assimilare le modalità di X a punti di una retta, e la diversità di due modalità alla distanza dei due punti corrispondenti. Perciò la « diversità » fra due modalità è qui caratterizzata in modo alquanto diverso dall'usuale: comunemente, infatti, la diversità fra due modalità si considera soltanto in valore assoluto, non facendosi, ad es., distinzione fra la diversità dal *do* al *mi* successivo, e la diversità da quel *mi* al *do* precedente. La diversità di due modalità nel senso sopra caratterizzato è, invece, la diversità in valore assoluto con l'aggiunta di un attributo: positiva o negativa; ma da tale aggiunta non si può prescindere se le serie rettilinee si vogliono sottoporre a calcolo. Ciò non significa, d'altronde, che in alcune investigazioni non sia necessario o conveniente considerare le sole diversità assolute delle modalità; e perciò si tenga presente che se A, B, C, *si seguono*, in un ordine qualunque, sarà:

$$|AC| = |AB| + |BC|; \; |AC| > |AB|; \; |AC| > |BC|$$

Si osservi ancora che l'essere una diversità maggiore di un'altra nel senso sopra fissato non significa che la prima sia maggiore della seconda nel senso usuale che è caratterizzato dalle relazioni ora scritte.

Resta ancora a vedersi se e come si possa giudicare che la diversità da A a B è uguale a quella da C a D. Se il carattere qualitativo X è associabile ad un carattere quantitativo X' le cui modalità $A', B', C', D', \ldots$ corrispondono rispettivamente alle $A, B, C, D, \ldots$ ed è $B' - A' = D' - C'$, allora viene naturale di ammettere che anche la diversità da A a B è uguale a quella da C a D. Per esempio se $A, B, C, D, \ldots$ sono i gradi di una gerarchia ed $A', B', C', D', \ldots$ sono gli stipendi corrispondenti, supposto che sia $B' - A' = D' - C'$, si giudica (anche usualmente) che la diversità da A a B sia uguale a quella da C a D. Sorge una difficoltà qualora si connettano ad X diversi caratteri quantitativi X' X'', ... le cui modalità A', B', C',

$D', \ldots$ ed A'', B'', C'', $D'', \ldots$, siano quelle che corrispondono rispettivamente ad A, B, C, D; infatti ammettendo che sia $B' - A' = D' - C'$, potrà non essere sempre $B'' - A'' = D'' - C''$: ed allora non si potrà, in generale, dare un criterio per decidere quale fra tali caratteri debba preferirsi come strumento di paragone delle diversità fra le modalità di X. Così, nella successione dei colori di uno spettro, si può assumere come criterio di equidiversità dal colore A a B e dal colore C a D il fatto che sullo spettro stesso la distanza lineare da A a B (o fra le mediane delle zone nominate A e B) sia uguale a quella da C a D; ma si potrebbe anche assumere l'altro fatto che le differenze dei numeri di vibrazioni corrispondenti ad A e B e corrispondenti a C e D siano uguali. Coppie di modalità equidiverse secondo un criterio di valutazione, non lo saranno secondo l'altro criterio. In casi analoghi a questi può accadere che un esame critico della questione ci additi quello fra i diversi caratteri quantitativi connessi ad X, che deve essere preferito al fine di valutare le diversità fra le modalità di tale carattere X; ma, a togliere qualsiasi ambiguità, si dovrà naturalmente dichiarare quale sia il carattere quantitativo prescelto.

Se nessuno dei caratteri quantitativi connessi ad X abbia titolo di preferenza rispetto agli altri, o se non intervengano affatto caratteri quantitativi e talora anche in altri casi, è necessario ricorrere ad una semplice « stima » delle diversità intercedenti fra le varie modalità di X.

Ora, a questo proposito, è opportuno rilevare: che, in primo luogo, anche la stima è un mezzo, sia pure rudimentale, di misura; e, secondariamente, che tale mezzo si deve effettivamente impiegare anche nelle più scrupolose misurazioni pratiche dei caratteri quantitativi, quando si cerchi di ottenere una approssimazione che vada oltre l'unità di misura adoperata. Così in tutti i casi di applicazione del nonio interviene una stima ad occhio nell'apprezzare quali sono le divisioni dei due lembi più vicine fra di loro; quando si fa uso di un teodolite a microscopî si ricorre una prima volta ad una stima nel portare i fili del micrometro a comprendere meglio che sia possibile un tratto della divisione sul cerchio, e una seconda volta nell'apprezzare (ordinariamente in decimi) la frazione di una divisione della testa di vite micrometrica. Nell'appulso del sole al contatto o al distacco dei diversi fili del reticolo del cannocchiale col quale viene osservato il passaggio di quell'astro al meridiano, interviene un apprezzamento che non può generalmente essere con-

trollato e di cui si cerca di eliminare l'errore con la simmetria delle osservazioni o col fare in modo che l'apprezzamento applicato ad altre osservazioni, poste in relazione con le prime, conduca ad un errore contrario. Talora è conveniente apprezzare ad orecchio i decimi del minuto secondo battuti da un cronometro. Si può dunque veramente dire che, in qualunque più scrupolosa misurazione quantitativa, interviene un apprezzamento soggettivo analogo a quello che ci guida nel giudizio delle diversità fra i vari gradi di un carattere qualitativo.

Sola differenza fra la stima applicata alla valutazione delle quantità e quella che si deve adoperare per la valutazione delle diversità fra modalità qualitative è questa: che la prima è generalmente preceduta da un certo numero di esperienze consistenti in concrete misurazioni, non trascendenti l'unità di misura adoperata, mentre la seconda non può essere preceduta da un tirocinio pratico analogo a quello (almeno se il carattere qualitativo non si connette ad uno quantitativo). Ma non si può negare la possibilità di eseguire delle stime anche nel campo qualitativo, tanto è vero che esse appartengono alla pratica comune.

Fra due suoni di diversa altezza che possono essere emessi da una corda più o meno tesa, un orecchio esercitato potrà dire se passi o no lo stesso intervallo musicale che fra due altri suoni emessi dallo stesso strumento (1).

Fra le varie tonalità ottenute dai diversi miscugli di due colori, un pittore potrà facilmente costituire una graduazione in cui ciascuna tonalità abbia dalla successiva una diversità costante. Un maestro provetto saprà con buona approssimazione apprezzare se il progresso compiuto in un certo tempo da un suo allievo sia o no, pari a quello compiuto da un altro allievo.

(1) Evidentemente, quando si dice che due intervalli musicali sono uguali, si tende a considerarli come due differenze uguali, tanto è vero che si dice essere la differenza di tono fra il *do* e il *do diesis* uguale a quella fra il *do diesis* ed il *re* (scala temperata). Ora un tale apprezzamento di differenze uguali, è forse connaturato alla struttura dell'orecchio umano; ma se si volesse ricorrere alla misura di un carattere quantitativo connesso al carattere qualitativo « altezza del suono » emesso da quella corda, sarebbe naturale riferirsi al numero di vibrazioni corrispondente a ciascuna modalità, e allora si vedrebbe che quelle che il nostro orecchio apprezza come *uguali differenze* di tono sono in effetto *uguali rapporti*, fra i corrispondenti numeri di vibrazioni. Ecco dunque che, anche quando il carattere qualitativo X si connetta ad un quantitativo, non sempre conviene attenersi alla misura delle modalità di questo ultimo, per definire le uguali diversità fra le modalità di X.

Non si fa, e non si può fare, in questi casi, una vera esperienza di confronto, ma ci si deve appagare di un apprezzamento sommario o di una stima. E se in questo processo entra in buona parte l'elemento personale, non è tuttavia da escludersi la possibilità di affinare questo senso di giudizio in modo che i risultati delle stime siano attendibili e concordanti qualora esse vengano eseguite da diverse persone.

Comunque, o che si faccia ricorso alla misurazione di un carattere quantitativo connesso al carattere qualitativo X, o che si proceda per via di una semplice stima, *ammetteremo di saper dire, in tutti i casi pratici che ci si presenteranno se la diversità dalla modalità* A *a* B *è (o non è) uguale a quella da* C *a* D*; cosicchè dovremo intendere che il concetto di uguaglianza sia applicabile nei confronti delle diversità fra le modalità di* X.

Corrispondenza fra modalità di un carattere rettilineo e numeri

7\. — È ora breve il passo per giungere a far corrispondere ad ognuna delle modalità di un carattere qualitativo rettilineo X un numero reale, in modo che a numeri equidifferenti corrispondano modalità equidiverse.

Intanto, se al carattere X conviene connettere un certo carattere quantitativo X' è ovvio che i numeri da far corrispondere alle modalità di X saranno le misure delle corrispondenti modalità di X' rispetto ad una unità arbitraria. Altrimenti si procederà come segue:

a) Se le successive modalità di X hanno ciascuna dalla precedente una diversità costante H (se cioè la serie è di quelle che sono state dette *equispaziate*, mentre quella diversità costante è stata detta *ragione* (1) per evidente analogia a ciò che accade in una progressione aritmetica) allora, fatto corrispondere ad una modalità M qualunque di X un numero reale m, e alla diversità costante il numero reale h, le modalità successive ad M avranno come corrispondenti i numeri $m + h$, $m + 2h, \ldots$ e le precedenti i numeri $m - h$, $m - 2h, \ldots$ Quella modalità M si potrà dire *origine*. Nel modo più semplice si potrà assumere $h = 1$ (cioè prendere per unità di misura delle diversità quella che intercede fra due modalità suc-

(1) C. Gini. *Variabilità e mutabilità*, già cit.

cessive), e fare corrispondere alla prima modalità A un numero intero a. Comunque, una qualunque delle progressioni:

$$\ldots m - 2h,\ m - h,\ m,\ m + h, \ldots$$
$$a,\ a + 1,\ a + 2, \ldots$$

non rappresenterà coi suoi termini la misura delle modalità di X ma sarà, come vedremo, sostituibile alla successione di tali modalità, nello studio numerico di alcuni caratteri di una serie data in funzione di X.

Al contrario, i numeri $0, h, 2h, 3h, \ldots$ e $-h, -2h, \ldots$ si considerano effettivamente come *misure* delle diversità fra le modalità, e tali misure si potranno dire *scostamenti* delle modalità fra di loro.

b) Se, invece, le successive modalità di X non hanno fra di loro una diversità costante, ammettiamo di sapere, almeno idealmente, intercalare fra le modalità effettive altre modalità fittizie, in modo da ottenere una successione di modalità equispaziate.

Precisamente (Cfr. n. 5), se X è un carattere discreto si tratterà di eseguire una vera intercalazione di nuove modalità; se X è continuo bisognerà operare una nuova distinzione in modalità opportunamente nominate ed equispaziate, tale però da comprendere anche le primitive modalità.

Alle nuove modalità si potranno poi, come prima, far corrispondere i termini successivi di una progressione aritmetica, cosicchè, eliminati, se occorre, gli elementi fittizi, alle modalità effettive di X corrisponderanno, in generale, termini non consecutivi di quella progressione, sui quali si potrà poi operare per la trattazione numerica della serie data.

Notiamo, circa la possibilità di eseguire quella intercalazione in modo da ottenere una successione equispaziata di modalità, che non vogliamo neppure lontanamente adombrare a relazioni che potrebbero dirsi di commensurabilità fra le diversità delle modalità di X; ma teniamo presente che: sia che queste modalità costituiscano un insieme discreto, cioè non ammettano gradazioni di passaggio dall'una all'altra, sia che esse si possano considerare come facenti parte di un insieme praticamente continuo, accadrà sempre che le loro diversità, per quella stessa indeterminatezza che abbiamo visto essere insita nell'apprezzamento che se ne deve eseguire, si potranno esprimere come multipli interi di una diversità costante H, il che equivale appunto al sapere eseguire la detta intercalazione.

Tutto ciò equivale altresì ad ammettere che si sappia attribuire un significato a locuzioni come queste: « la diversità fra A e B è quintupla di quella fra P e Q », oppure: « quella fra C e D è $\frac{1}{3}$ di quella fra R ed S », o infine: « quella fra E ed F è $\frac{4}{7}$ di quella fra T e V » (1)

Applicazione dei procedimenti di calcolo alle serie rettilinee

8. — Si tratta, ora, di vedere come possa impiegarsi il sistema x di numeri che si è fatto corrispondere alle modalità di X, nello studio della serie che si suppone data in funzione di X.

Notiamo, intanto, che la corrispondenza posta è biunivoca e cioè che ad ogni modalità di X corrisponde un solo numero del sistema x, e viceversa.

(1) La circostanza che l'apprezzamento della diversità fra due modalità di un carattere qualitativo X si effettua di solito mediante una stima, fa sì che praticamente tali diversità risultano espresse da multipli assai semplici di un numero h, o se si vuole, da numeri interi assai piccoli. Anche quando il carattere X sia continuo, ma lo si sia distinto in modalità opportunamente nominate, queste si considerano poi come elementi di un insieme discontinuo, cioè come se ciascuna di tali modalità non presentasse alcuna gradazione (Cfr. Gini, *Var. e Mut.*, pag. 116). Ma, evidentemente, tale assimilazione delle modalità in cui praticamente si distingue un carattere continuo, a quelle di un carattere discreto ha soltanto scopo semplificativo. Del resto essa si opera, talora, anche per caratteri quantitativi.

Se un insieme di redditieri si distribuisce in classi parziali a seconda del reddito 0 — 1000; 1000 — 2000; 2000 — 3000;... queste classi parziali, che possiamo denotare come 1ª, 2ª, 3ª, ecc., formano un aggregato discreto e la differenza di reddito fra due redditieri appartenenti a classi consecutive, si presume essere genericamente di lire 1000. Ciò equivale ad ammettere che, entro ciascuna classe, i redditieri si distribuiscano in modo tale che i redditi medî corrispondenti alle varie classi differiscano successivamente di lire 1000; oppure anche ad ammettere che la differenza media fra redditi appartenenti a classi successive sia di lire 1000. Non diversamente si ammette che, distribuita una popolazione per classi annuali di età, la differenza media di età fra due persone appartenenti a classi consecutive sia di un anno, ecc.

Si tratta, in sostanza, di una espressione che sintetizza la scissione dell'insieme continuo in classi parziali, ma che ha significato rigoroso soltanto se le distribuzioni che si considerano sono uniformi entro ciascun intervallo.

Si abbiano sopra una retta due segmenti consecutivi di lunghezze a e b, e presa come origine O l'estremo del primo segmento, non comune col secondo, sia x l'ascissa

In secondo luogo osserviamo (a somiglianza di ciò che accade per i caratteri quantitativi) che *anche un numero non appartenente a quel sistema può, almeno in molti casi, assumere una significazione concreta, come rappresentante di una modalità di* X *diversa da quelle considerate.* Naturalmente è qui necessario distinguere il caso che X sia discreto da quello che sia continuo: nel primo si dovranno intercalare convenientemente nuove modalità, reali o fittizie, fra le date; nell'altro bisognerà suddividere l'intervallo di mutabilità di X in un diverso numero di modalità-classi.

Tuttavia, tanto per caratteri discreti che continui, *il processo di intercalazione di nuove modalità o di suddivisione dell'intervallo*

di un punto generico del primo segmento e ξ quella di un punto del secondo. La somma delle distanze di un punto generico (x) del primo segmento dai punti del secondo è:

$$\int_{a}^{a+b} (\xi - x)\, d\xi$$

e perciò la distanza media di quel primo punto dai secondi è:

$$\frac{1}{b}\int_{a}^{a+b} (\xi - x)\, d\xi = \left(a + \frac{b}{2}\right) - x.$$

Ne segue che la distanza media di un punto del primo segmento da un punto del secondo sarà:

$$\frac{1}{a}\int_{0}^{a} \left\{ \left(a + \frac{b}{2}\right) - x \right\} dx = \frac{a+b}{2}$$

e, se $b = a$, tale distanza media si riduce semplicemente ad a.

Per questa ragione appunto, diviso l'intervallo di variazione di un carattere quantitativo in intervalli uguali, la differenza fra due classi successive si riguarda uguale all'ampiezza di ciascun intervallo parziale; mentre se gli intervalli non sono uguali, ed a e b sono le ampiezze di due di essi consecutivi, la differenza fra le due classi corrispondenti si riguarda uguale ad $(a + b)/2$. Ma, ripetiamo, questi apprezzamenti sommarî della differenza fra due classi sono soltanto approssimativi, perchè le conclusioni ottenute pei punti di una retta si possono rigorosamente estendere ad una seriazione soltanto se entro ciascun intervallo la distribuzione sia uniforme.

Ciò si potrebbe ripetere evidentemente anche per caratteri qualitativi rettilinei ed ha motivo di frequente applicazione per le semplificazioni che ne possono conseguire. Cfr. in proposito Gini, *Variabilità e Mutabilità*, già cit., paragr. 73, pag. 130 ed anche: *Il concetto di « transvariazione » e le sue prime applicazioni*, in « Studi di Economia, Finanza e Statistica », dello stesso Autore.

di mutabilità potrebbe teoricamente proseguire fino al limite della percezione di una diversità fra due modalità, ma converrà praticamente arrestarlo anche prima.

Così, se alle ricompense al valor militare (carattere discreto a modalità non equispaziate)

encomio solenne, med. di bronzo, med. d'argento, med. d'oro

si fanno corrispondere i numeri

50 100 250 800

e viceversa, al numero 200 sarà ragionevole far corrispondere una modalità (fittizia e non particolarmente nominata) compresa fra la medaglia di bronzo e quella d'argento e più prossima a questa che non all'altra; al numero 75 che ha uguali differenze assolute da 50 e da 100, corrisponderà una modalità compresa fra l'encomio solenne e la medaglia di bronzo ed ugualmente diversa da queste due.

Se i gradi di una gerarchia

$$A, B, C, D, \ldots$$

si fanno corrispondere ai numeri (che potrebbero essere i rispettivi stipendi)

10.000, 15,000, 20.000, 25.000,...

e viceversa, al numero 18.000 si potrebbe far corrispondere « il grado B con una conveniente anzianità », ecc.

Se infine, tenendo conto dell'*intervallo* musicale fra le successive note della scala musicale in tono maggiore (la quale non è equispaziata), si ponesse la corrispondenza

do	*re*	*mi*	*fa*	*sol*	*la*	*si*
2	4	6	7	9	11	13

(assumendo come termine di paragone delle diversità l'intervallo di mezzo tono), al numero 5 corrisponderebbe nella scala temperata il *re diesis* o *mi bemolle*, al numero 3,25 una nota (che non ha una particolare denominazione e che non dà luogo ad accordi con altre note della stessa scala, ma) che sarebbe compresa fra il *do diesis* e il *re* e che si potrebbe produrre sperimentalmente, ecc.

Ciò posto, per analogia a ciò che accade per le seriazioni (date, dunque, in funzione di caratteri quantitativi) si assumerà questo principio generale: *Data una serie rettilinea in funzione del carattere qualitativo* X, *e fatti corrispondere alle modalità* X_i *di* X *dei nu-*

meri x_i, *nel modo già dichiarato, si eseguiscano su tali numeri le stesse operazioni che applicate alle misure di caratteri quantitativi* (o a trasformate lineari di tali misure) *fornivano la media aritmetica o la mediana o la moda o gli ordinari indici di variabilità delle seriazioni basate su questi caratteri. Se i risultati di quelle operazioni conservano un significato rispetto alla serie data, tali risultati si diranno costituire la media aritmetica o la mediana o la moda o rispettivamente gli indici di* « *mutabilità* » *della serie data.*

Ora, per quanto concerne la *media aritmetica,* basterà dunque trovare la media aritmetica a dei numeri x_i, attribuendo loro come pesi i valori Y_i del carattere quantitativo Y che si considera funzione di X nella data serie, ossia $a = \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i}$. La modalità A di X che, in base alle osservazioni premesse si dovrà fare corrispondere ad a sarà, per definizione, la media aritmetica A di X nella data serie (1). Naturalmente – poichè per una seriazione la media aritmetica si può anche definire come modalità per cui è nulla la somma algebrica degli scostamenti pesati, o come modalità per cui è minima la somma dei quadrati pesati degli scostamenti, o come baricentro dei punti rappresentativi delle modalità, essi pure opportunamente pesati – così anche per le serie rettilinee si potrà indifferentemente caratterizzare la media aritmetica con una qualunque di queste definizioni.

La *mediana* sarà costituita da quella o da una di quelle modalità X_i di X tali che le somme dei valori di Y corrispondenti alle modalità di X precedenti e non precedenti X_i siano rispettivamente minore e maggiore od uguale alla semisomma di tutti i valori di Y.

(1) Per esempio nella valutazione del merito X dei giovani di una scolaresca siano impiegati i termini insufficiente, sufficiente, buono, ottimo; si domanda quale sia il merito medio di una scolaresca di 30 giovani, dei quali 6 insufficienti, 12 sufficienti, 8 buoni, 4 ottimi. Ammesso che questi gradi di merito debbano considerarsi come equispaziati, e fatti ad essi corrispondere i termini di una progressione aritmetica, per es.: 1, 2, 3, 4, il merito medio cercato sarà espresso dal numero 70/30 = 2 1/3, e quindi compreso fra le modalità « sufficiente » e « buono ».

Se nella successione delle modalità di X si intercalano altre modalità equispaziate (e precisamente due in ognuno degli attuali intervalli) al numero 2 1/3 corrisponderà una di queste nuove modalità. Se, invece, tale intercalazione non si vuole o non si può eseguire (perchè trascende la nostra facoltà di giudizio nell'apprezzamento delle diversità fra le modalità di X) allora, praticamente, al numero 2 1/3 corrisponderà la modalità rappresentata da quello dei numeri 1, 2, 3, 4 che gli è più prossimo, cioè « sufficiente ».

Moda o *norma* sarà una qualunque delle modalità di X a cui corrisponde il massimo assoluto o uno di massimi relativi di Y

Infine, gli indici di variabilità del sistema di numeri x, ponderati come si è detto per la media aritmetica, si diranno costituire gli *indici di mutabilità* della serie data, sui quali non ci soffermeremo, rimandando senz'altro il lettore alla trattazione sistematica che ne è stata fatta altrove (1).

Vogliamo soltanto osservare, a tale proposito, che, come nel caso delle seriazioni gli ordinari indici assoluti di variabilità rappresentano quantità omogenee con i caratteri quantitativi in funzione dei quali esse sono state costituite, così, nel caso delle serie rettilinee, un indice assoluto di mutabilità deve pensarsi espresso in termini di quella « diversità » che si è presa come unità nella misura delle diversità fra le varie modalità del carattere X.

Concludendo, per un carattere qualitativo X che dia luogo ad una serie rettilinea non si può stabilire una teoria della misura, come si fa per caratteri quantitativi; è tuttavia possibile, talora considerando un carattere quantitativo che si connette ad X, o, più spesso, procedendo per mezzo di stime, valutare le diversità fra le successive modalità di X, e far quindi corrispondere un numero a ciascuna di queste modalità, partendo da un numero arbitrario fatto corrispondere alla prima di esse.

Sui numeri così ottenuti (che non sono misure delle modalità) si può operare come su quelli che misurano le modalità di un carattere quantitativo ed ottenere dei risultati che rappresentano la media aritmetica, la mediana, la moda e gli ordinari indici di mutabilità della serie.

Brevemente diremo, dunque, come già altrove è stato affermato (2) che le serie rettilinee si possono matematicamente investigare, *ai fini detti*, nello stesso modo delle seriazioni.

(1) C. GINI. *Variab. e Mutab.*, già cit.

(2) C GINI. *Di una estensione...* già cit., pag. 408. Come applicazione di tale criterio vedasi, p. es. per serie rettilinee equispaziate. C. GINI, *Variabilità e mutabilità*, già cit. § 68; e per serie rettilinee qualunque: G. ZINGALI. *Cenni statistici sulle medaglie al valor militare*. « La Riforma Sociale », XXIX, 1918. Anche G. TAGLIACARNE, *Contributi e comportamenti delle Regioni d'Italia in guerra*, « Metron », Vol II, 1923, trasforma il carattere qualitativo delle medaglie al valor militare in un carattere quantitativo, in base all'assegno annuo, che è di lire 800 per le medaglie d'oro, di lire 250 per quelle d'argento, e di lire 100 per quelle di bronzo.

Completiamo questo cenno sulle serie rettilinee osservando che, se i valori Y non sono tutti positivi, la media ponderata $a = \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i}$ può rappresentare una modalità di X non compresa fra la prima e l'utima data, ma esterna al loro intervallo. Se, pertanto, si volesse riservare la denominazione di «modalità medie» trà più date soltanto a quelle comprese fra la prima e l'ultima data, la modalità corrispondente ad a nel caso detto non si dovrebbe dire una «modalità media» fra le date. Ma per generalità di linguaggio e soprattutto per la circostanza che anche quando la a rappresenti una modalità cosiffatta permangono in essa alcune proprietà formali delle medie aritmetiche ponderate, è ben lecito mantenerle, anche allora, tale denominazione. Avremo più oltre occasione di ritornare su questo concetto.

III. — SERIE CICLICHE

Misura della diversità fra le modalità di un carattere ciclico

9. — Vediamo come il principio di misurare non le modalità di un carattere qualitativo ma le loro diversità, già utilmente sfruttato per le serie rettilinee, possa avere impiego anche per le serie cicliche.

Se X è un carattere ciclico, le sue modalità si seguiranno in due sensi od ordini naturali contrarî, dei quali fisseremo uno come positivo e l'altro come negativo. Essendo A e B due modalità qualunque, accadrà che – a differenza di ciò che accadeva pei caratteri rettilinei – tanto nel senso positivo come in quello negativo ad A seguirà B. Perciò la diversità da A a B potrà essere valutata in ciascuno dei due sensi, e altrettanto si potrà dire per la diversità da B ad A. Mentre dunque per le serie rettilinee esiste una sola diversità fra due modalità delle quali una sia considerata come prima, per le serie cicliche ne esistono due. Indicheremo le due diversità da A a B, quella nel senso positivo e quella nel senso negativo, rispettivamente con $(\overline{AB})$ e $(\underline{AB})$; e similmente con $(\overline{BA})$ e $(\underline{BA})$ le due diversità da B ad A. La diversità da A a B in

un senso e quella da B ad A in senso contrario si diranno opposte fra loro, e si potrà scrivere:

$$(1) \qquad \begin{cases} (\overline{AB}) = -(\underline{BA}), \; (\underline{AB}) = -(\overline{BA}) \\ (\overline{BA}) = -(\underline{AB}), \; (\underline{BA}) = -(\overline{AB}) \end{cases}$$

Si dirà altresì che le due diversità $(\overline{AB})$ e $(\underline{BA})$ hanno lo stesso valore assoluto, il quale si indicherà indifferentemente con $|\overline{AB}|$ e $|\underline{BA}|$, cosicchè $|\overline{AB}| = |\underline{BA}|$; e similmente si avrà $|\underline{AB}| = |\overline{BA}|$.

Se A, B, C, sono modalità qualunque, esiste sempre un *certo ordine* in cui esse saranno consecutive, ed allora sarà naturale di dire che, *nello stesso ordine*, la diversità da A a C è la somma delle diversità da A a B e da B a C, e si scriverà $(\overline{AC}) = (\overline{AB}) + (\overline{BC})$ oppure $(\underline{AC}) = (\underline{AB}) + (\underline{BC})$ secondo che quell'ordine sia il positivo o il negativo.

Sussisterà dunque sempre una delle due relazioni:
$|\overline{AC}| = |\overline{AB}| + |\overline{BC}|$ oppure $|\underline{AC}| = |\underline{AB}| + |\underline{BC}|$ e quindi anche:
$|\overline{AC}| > |\overline{AB}|$, $|\overline{AC}| > |\overline{BC}|$ oppure $|\underline{AC}| > |\underline{AB}|$, $|\underline{AC}| > |\underline{BC}|$.

Se, per esempio, $(\overline{AC}) = (\overline{AB}) + (\overline{BC})$ diremo anche che $(\overline{AB}) = (\overline{AC}) - (\overline{BC})$ da cui:

$$(2) \qquad (\overline{AB}) = (\underline{CB}) - (\underline{CA}).$$

D'altra parte, in quello stesso ordine, anche le modalità C, A, B sono consecutive, e perciò $(\overline{CB}) = (\overline{CA}) + (\overline{AB})$ da cui:

$$(3) \qquad (\overline{AB}) = (\overline{CB}) - (\overline{CA})$$

Avvicinando le 2) e 3) si conclude che *in quello stesso verso* in cui le modalità A, B, C sono consecutive, la diversità da A a B è la differenza fra le diversità dalla terza modalità alla seconda e la diversità dalla terza alla prima, essendo prese arbitrariamente queste due diversità entrambe nel senso positivo o entrambe nel senso negativo.

Se invece si considera la diversità di due modalità A e B nel verso opposto a quello in cui A, B, C sono consecutive (se p. es. A, B, C sono consecutive nel verso negativo e si considera la $(\overline{AB})$, si trova: $(\overline{AB}) = (\overline{AC}) + (\overline{CB})$ cioè:

$$(4) \qquad (\overline{AB}) = (\overline{CB}) - (\underline{CA})$$

Dunque, *nel verso contrario* a quello in cui A, B e C sono consecutive, la diversità da A a B è la differenza di quelle da C a B e

da C ad A, essendo presa quella da C a B nello stesso verso della diversità da A a B, e l'ultima nel verso contrario.

Essendo A e B due modalità qualunque di X, diremo « ampiezza del ciclo di mutabilità di X » la somma delle diversità da A a B e da B ad A, considerate entrambe nello stesso senso. Se, senza tralasciarne alcuna, tutte le modalità di X che si incontrano dalla A in poi sono in un certo senso $A, B, C, \ldots M, N$, la somma delle diversità da A a B, da B a $C \ldots$ da M ad N, da N ad A, risulterà uguale all'ampiezza del ciclo.

Anche qui, senza ripetere il già detto per le serie rettilinee, ammetteremo di saper dire in tutti i casi pratici che ci si presenteranno (ed a seguito della misura di un carattere quantitativo eventualmente collegato ad X, o come risultato di una semplice stima) se in un certo senso, la diversità da A a B è o non è uguale a quella da C a D.

Se, nel senso positivo, la diversità da A a K è uguale a quella da K ad A, cioè se $(\overline{AK}) = (\overline{KA})$ si dice che le modalità A e K sono *opposte* e ciascuna di quelle diversità si dice essere di mezzo ciclo positivo.

Se ne deduce che $-(\underline{KA}) = -(\underline{AK})$ e quindi che $(\underline{AK}) = (\underline{KA})$; ossia le due diversità sono anche uguali nel senso negativo.

Se, in un certo senso, la diversità da A ad L non è uguale a quella da L ad A, una di esse sarà minore e l'altra maggiore di mezzo ciclo, in valore assoluto.

Se, in un certo senso ed in valore assoluto la diversità da A ad L è minore, e quella da L ad A è maggiore di mezzo ciclo, nel senso contrario la diversità da A ad L sarà maggiore e quella da L ad A sarà minore di mezzo ciclo in valore assoluto. Quindi, in valore assoluto, le due diversità da A ad L sono entrambe uguali a mezzo ciclo, oppure una di esse è minore di mezzo ciclo.

Infine, se invece di considerare due modalità distinte, si considera due volte una stessa modalità, A ed A, le due diversità che abbiamo detto esistere in un medesimo senso si riducono l'una a nulla e l'altra a un ciclo.

Le convenzioni ed osservazioni precedenti sono evidentemente suggerite dall'assimilare le modalità di un carattere ciclico a punti di una circonferenza e le due possibili diversità dalla modalità A alla B ai due archi complementari che vanno dal punto immagine di A al punto immagine di B. Se ora osserviamo tutte le proposizioni testè enunciate, vediamo che esse sono esplicitamente subordi-

nate alla fissazione del senso in cui si considera la diversità da una ad un'altra modalità del ciclo, non bastando sapere quale di due modalità si consideri come prima per determinare quella diversità, come invece accadeva per i caratteri rettilinei.

Pertanto, se la teoria delle serie cicliche si vuol basare su criterî quanto più sia possibile analoghi a quelli posti per le serie rettilinee si prospetta anzitutto la convenienza di togliere ambiguità di significato alla locuzione « diversità da A a B »; e ciò si ottiene mediante la seguente convenzione:

Se in un certo senso la diversità da A *a* B *non è uguale in valore assoluto a quella in senso contrario, diremo « diversità base da* A *a* B » *e indicheremo con* (AB) *quella delle due che ha minore valore assoluto cioè che è minore di mezzo ciclo in valore assoluto; se invece le diversità da* A *a* B *sono in valore assoluto uguali nei due sensi, cioè se* A *e* B *sono modalità opposte, diversità base sarà indifferentemente una delle due, cioè mezzo ciclo positivo o mezzo ciclo negativo* (1).

A proposito di tale convenzione due rilievi sono peraltro necessarî. Anzitutto essa non riesce ad eliminare un caso di ambiguità che si presenta quando le A e B sono opposte. In secondo luogo essa non consente di definire la disuguaglianza delle diversità fra modalità mediante l'addizione delle diversità stesse. Difatti, se A, B, C sono modalità consecutive in un certo ordine e la diversità da A a C nello stesso ordine è in valore assoluto, maggiore di mezzo ciclo, mentre la diversità da A a B e quella da B a C sono, ciascuna, minori di mezzo ciclo, la diversità base da A a C non è la somma delle diversità basi da A a B e da B a C; e non si può quindi neanche dire che la diversità base da A a C sia in valore assoluto maggiore di quella da A a B, e di quella da B a C.

Nonostante ciò, se si vogliono sottoporre a calcolo le serie cicliche, basandosi sulla diversità delle modalità come si è fatto per le serie rettilinee, è preferibile rinunciare alla conservazione della solita definizione della disuguaglianza, piuttosto che alla unicità di significato della locuzione « diversità dalla modalità A alla modalità B ». Esigere l'una e l'altra condizione sarebbe impossibile, per la natura stessa dei caratteri ciclici: e perciò la rinuncia deve cadere su quella delle due che si presenta come meno essenziale negli ulte-

(1) Salvo l'aggiunta dell'aggettivo « base » *v.* per questa convenzione: C. Gini. *Variab. e Mutab.* già cit.

riori sviluppi. In ogni modo, è questo il punto dal quale si delinea una delle difficoltà che sovrastano alla trattazione matematica delle serie cicliche.

Notiamo, tuttavia, che se le diversità base delle varie modalità si misurano tutte rispetto ad una stessa modalità O ed A, B sono modalità qualunque appartenenti insieme con O ad uno stesso emiciclo, si ha, come per i caratteri rettilinei:

$$(AB) = (OB) - (OA)$$

D'ora in poi, salvo avviso contrario, intenderemo per diversità da A a B la diversità base (AB).

Corrispondenza fra modalità di un carattere ciclico e numeri

10. — Ciò posto, passiamo a mostrare come la considerazione delle diversità (base) fra le modalità di un carattere qualitativo ciclico, ci conduca a far corrispondere ad ognuna di tali modalità un numero reale. La misura della diversità base dalla modalità A alla B diremo *scostamento* da A a B o di B rispetto ad A.

a) I caratteri ciclici più spesso considerati sono quelli in cui la diversità base fra le successive modalità e fra l'ultima e la prima è costantemente uguale ad H. Le serie che ne risultano si possono dire « equispaziate a ragione H » (1). Si dice anche che la diversità espressa da H è di un grado. Fatto allora corrispondere alla diversità costante H un numero reale h, l'ampiezza del ciclo sarà rappresentata da sh, essendo s il numero delle modalità; o più semplicemente dal numero intero s, se si prende $h = 1$.

In questa ipotesi, se r è la diversità base da A ad L, la diversità da A ad L nel significato generale (Cfr. n. 9) e nel senso contrario a quello della diversità base sarà $r \mp s$ secondo che sia $r \gtrless 0$. Le due diversità da A ad L in senso generale, sono dunque espresse da numeri congrui rispetto ad s preso come modulo.

Quando il numero delle modalità del ciclo è pari, ognuna di esse ne ammette una opposta; quando sia dispari si dicono, per estensione, opposte a ciascuna modalità quelle due che hanno dalla prima la massima diversità in valore assoluto.

(1) C. Gini. *Variab. e Mutab.*, già cit.

Allora, se il numero s delle modalità è dispari, $s = 2k + 1$, presa come unità di diversità quella fra due modalità consecutive, presa come prima modalità od origine una qualsiasi di esse M e fattole corrispondere il numero m, alle k modalità successive nell'ordine positivo, si faranno corrispondere i numeri $m + 1$, $m + 2 \dots m + k$ ed alle k precedenti i numeri $m - 1$, $m - 2, \dots m - k$; se invece s è pari, $s = 2k$, tutto si potrà ripetere nello stesso modo, salvo che *alla modalità opposta a quella iniziale corrisponderanno simultaneamente i due numeri* $m \pm k$. Converrà dunque sempre distinguere, anche in ciò che segue, il caso di s pari da quello di s dispari, ma intanto concludiamo che alle modalità di un carattere qualitativo X che dia luogo ad una serie ciclica equispaziata, si possono far corrispondere termini consecutivi di una progressione aritmetica.

Osserviamo tuttavia che:

1°) Quei termini non possono riguardarsi come misure delle modalità; ma, se mai, come indicatori del posto che queste hanno nel ciclo rispetto alla modalità presa come iniziale.

2°) Pure essendo i termini stessi dedotti dalla considerazione delle diversità fra le modalità di X, *a differenza di ciò che accadeva per le serie rettilinee* non si può più dire che la diversità base da una ad un'altra modalità sia sempre espressa dalla differenza dei numeri corrispondenti. Per es. se ai giorni della settimana, che si possono considerare come le modalità di un carattere ciclico, si fanno corrispondere i seguenti numeri:

L	*Ma*	*Me*	*G*	*V*	*S*	*D*
—3	—2	—1	0	1	2	3

la diversità base da V a L, espressa dal numero 3, non è affatto la differenza dei numeri 1 e —3. Invece, la diversità base da L a Me è espressa dalla differenza $-1 - (-3) = 2$. Evidentemente se le due modalità e l'origine sono in uno stesso emiciclo, la diversità base sarà la differenza fra il numero corrispondente alla seconda e quello corrispondente alla prima modalità, così come si verificava per le serie rettilinee; se, invece, le due modalità e l'origine non sono nello stesso emiciclo la diversità base si avrà formando la differenza fra il numero corrispondente alla seconda e quello corrispondente alla prima e prendendo col segno cambiato il complemento ad s di questa differenza; e ciò perchè la diversità base ha senso contrario a quello che risulterebbe dal fare soltanto la differenza di quei numeri. In conclusione, se rispetto ad una origine,

i numeri corrispondenti alle modalità A e B sono a e b, la diversità base da A a B, sarà espressa da

$$(AB) = b - a$$

oppure da

$$(AB) = -(|b - a| - s)$$

(s numero delle modalità del carattere ciclico) secondo che A e B giacciono o non giacciono in uno stesso emiciclo insieme con l'origine.

3°) Ne segue che cambiando la modalità che si considera come prima o fondamentale, può cambiare la differenza in un dato ordine dei numeri corrispondenti a due certe modalità; così se ai giorni della settimana, prendendo come origine il sabato, si fanno corrispondere i seguenti numeri:

Me	*G*	*V*	*S*	*D*	*L*	*Ma*
—3	—2	—1	0	1	2	3

la differenza dei numeri corrispondenti ad L e S è $2 - 0 = 2$, mentre nella corrispondenza precedente era $-3 - 2 = -5$

In tutti i casi, se le due differenze non sono uguali, esse hanno segni contrari e valori assoluti complementari rispetto ad s, numero delle modalità del ciclo, ossia quelle differenze sono congrue rispetto al modulo s.

4°) Se avendo fatto corrispondere ad una modalità A, presa come iniziale, il numero m, corrisponde ad un'altra modalità L il numero $m + t = n$, presa quest'altra modalità L come iniziale e fattole ancora corrispondere il numero n, non si potrà dire che i numeri

$$\ldots n - 2,\; n - 1,\; n,\; n + 1,\; n + 2, \ldots$$

che se ne deducono come corrispondenti alle varie modalità del ciclo siano tutti quegli stessi che si erano avuti prendendo come modalità iniziale A e come numero corrispondente m, e cioè siano:

$$\ldots m - 1,\; m,\; m + 1, \ldots$$

Tuttavia, se ad una modalità P corrisponde rispetto all'origine A il numero p_1 e rispetto all'origine L il numero p_2, sarà $p_1 \equiv p_2$ (mod. s); e, viceversa, se $p_1 \equiv p_2$ (mod. s), ai numeri p_1 e p_2 corrisponderà una stessa modalità. Così nel ciclo dei giorni della settimana, preso l'inizio S e fattogli corrispondere lo zero, si avrà la corrispondenza:

Me	*G*	*V*	*S*	*D*	*L*	*Ma*
—3	—2	—1	0	1	2	3

Presa ora come iniziale L e fattovi corrispondere il 2, risulterà la corrispondenza

$$\begin{array}{ccccccc} Me & G & V & S & D & L & Ma \\ 4 & 5 & -1 & 0 & 1 & 2 & 3 \end{array}$$

Nelle due corrispondenze a Me e G corrispondono numeri diversi, ma è:

$$-3 \equiv 4 \pmod{7}, \quad \text{e} \quad -2 \equiv 5 \pmod{7}.$$

Questa osservazione si dovrà sempre tener presente in ciò che segue per quanto riguarda il significato delle notazioni, e cioè *X_p ed X_t rappresenteranno una stessa modalità di X, e similmente Y_p ed Y_t uno stesso valore di Y, se $p \equiv t$* (mod. s).

Da tutto ciò si vede che, pure avendosi la possibilità di far corrispondere alle modalità di un carattere ciclico equispaziato termini consecutivi di una progressione aritmetica, come accadeva per le modalità di un carattere rettilineo, pur nondimeno tale corrispondenza perde alcune delle proprietà che aveva in questo caso. Di qui una più difficile applicabilità delle operazioni di calcolo alle serie cicliche, rispetto alle rettilinee.

b) Se le varie modalità del ciclo non hanno ciascuna dalla seguente e l'ultima dalla prima una diversità costante, ammettiamo che, dal punto di vista pratico, sia possibile intercalare fra le modalità effettive altre modalità fittizie in modo da ottenere una serie ciclica equispaziata (1). Fatte corrispondere a tutte queste modalità altrettanti termini consecutivi di una progressione aritmetica, ed eliminate, se occorre, le modalità fittizie, alle modalità effettive verranno a corrispondere termini (in generale non consecutivi) di quella certa progressione; e su tali termini si potrebbero ripetere le stesse considerazioni fatte pel caso delle serie cicliche equispaziate.

c) Se, infine, il carattere X fosse continuo, converrebbe, come già sappiamo, spezzare il ciclo continuo in un insieme discreto di classi parziali o modalità-classi, in modo che queste varie modalità si potessero considerare come equispaziate; e si perverrebbe così ad esprimere le diversità delle varie modalità mediante numeri interi o mediante multipli di uno stesso numero h (2).

(1) Cfr. quanto si è detto in proposito al n. 7 per le serie rettilinee.

(2) Per esempio il ciclo settimanale può essere diviso in 7 modalità equispaziate (giorni della settimana), ma potrebbe anche essere diviso in 168 modalità equi-

Si tenga anche qui ben presente la circostanza che se al sistema delle modalità di X corrisponde un insieme di termini di una progressione aritmetica, o, più semplicemente, un insieme di numeri interi, può in taluni casi assumere una significazione concreta anche un numero non appartenente, ma compreso fra due termini di quell'insieme, come rappresentante di una modalità di X, non appartenente ad esso, ma ad un nuovo sistema costituito da un maggior numero di modalità. Questo si otterrà o per intercalazione di nuove modalità, reali o fittizie, se il carattere X è discreto, o per spezzamento del ciclo in un maggior numero di modalità-classi se X è continuo. La intercalazione o la suddivisione in nuove modalità si arresta, come pei caratteri rettilinei, appena cessa la facoltà di percepire una diversità fra due modalità.

Così, se il carattere ciclico (continuo) « direzione del vento » si considera come distinto in sole quattro modalità o quadranti N, E, S, W (riguardando come direzioni uguali tutte quelle comprese in uno stesso quadrante), e si pone la corrispondenza

S	*W*	*N*	*E*	*S*
—2	—1	0	1	2

anche il numero 1,5 si potrà considerare come rappresentante una modalità, purchè il dato carattere si distingua in 8 modalità od ottanti N, N-E, E, S-E, S, S-W, W, N-W (riguardando come direzioni uguali quelle comprese in uno stesso ottante), e si ponga

S	*S-W*	*W*	*N-W*	*N*	*N-E*	*E*	*S-E*	*S*
—2	—1,5	—1	—0,5	0	0,5	1	1,5	2

Se, invece, un carattere ciclico sia o convenga considerare come discreto, ad un numero compreso fra due del sistema dato si potrà far corrispondere una modalità fittizia convenientemente intercalata fra le modalità date; o, se non conviene eseguire tale intercalazione,

spaziate (ore della settimana) ed anche in qualunque altro numero di modalità, equispaziate oppure no. Le modalità del ciclo settimanale sarebbero dunque infinite: la divisione in classi parziali o modalità-classi è soltanto semplificativa, in quanto consente di riguardare come costante la diversità fra una modalità qualunque di una classe ed una modalità qualunque di un'altra classe. Così se il ciclo settimanale si divide in giorni nell'intento di classificare più eventi a seconda dei giorni della settimana, la diversità fra due eventi, uno accaduto in un istante qualunque di lunedì e l'altro in un istante qualunque di un martedì, si considererà sempre di un grado, ossia di un giorno. Si riveda a questo proposito, il n. 7-*b*), e la relativa nota a piè di pagina.

si dovrà fare corrispondere allo stesso numero la modalità indicata dal numero che gli è più prossimo fra i dati (Cfr. in proposito nota al n. 8). Naturalmente, una volta che una nuova modalità sia stata introdotta fra le date, si potranno anche determinare le diversità di queste rispetto alla nuova introdotta.

Applicazione del principio di conservazione delle leggi formali alla ricerca delle modalità medie di una serie ciclica

11. — Data una serie ciclica in funzione di un carattere X e fatte corrispondere le modalità di X a numeri convenienti, bisogna vedere se e come questi possano impiegarsi nella determinazione di modalità medie e di indici di mutabilità della serie.

A questo proposito si rilevi anzitutto che tutte le derivazioni logiche del concetto di *ordine naturale di successione* delle modalità di un carattere qualitativo *implicante l'esistenza di una prima e di un' ultima modalità* possono sussistere soltanto per le serie rettilinee. Così, se per estensione del concetto di valore medio, definito come un valore compreso tra il minimo ed il massimo di un insieme di numeri reali, si definisce come modalità media una qualunque modalità compresa tra la prima e l'ultima di un carattere qualitativo, evidentemente tale definizione sarà applicabile soltanto alle serie rettilinee.

Ma se, applicando il principio di conservazione delle leggi formali (Hankel) (1), si assumerà come definizione di una certa « modalità media » qualche sua proprietà caratteristica, *indipendente dalla esistenza di una prima e di un'ultima modalità*, potrà bene accadere

(1) Questo principio, di continua applicazione nell'analisi matematica, consiste in ciò: per una classe di enti (A), si definisca mediante (D) una certa operazione (O) la quale goda di certe proprietà caratteristiche (P), cosicchè nella classe (A) si possa sostituire alla (D) l'insieme delle proprietà (P) per definire la (O). Se la classe (A) viene ampliata, ottenendone una nuova classe (A') e in questa la definizione (D) perde significato, ma si può nondimeno definire una operazione, la quale in (A') goda delle proprietà (P) e si riduca ad (O) quando si applichi agli elementi di (A), *a tale operazione si attribuisce la stessa denominazione che si usa per la* (O). Tale principio si riferisce anche al risultato della operazione considerata, e noi lo applicheremo modificandolo nel senso che la nuova classe (A') possa anche non comprendere la (A).

che una tale definizione abbia senso anche per le serie cicliche (e per quelle sconnesse), come in altra occasione è stato già osservato (1).

Scopo principale della presente memoria è appunto quello di vedere se e come a questi tipi di serie si possano estendere i concetti delle medie consuete: media aritmetica, mediana, moda.

Dopo ciò si potranno definire quelle misure della mutabilità, nelle quali è necessario l'impiego di una modalità media. Infine vedremo se gli stessi concetti si possano generalizzare al caso di fenomeni a due o a più dimensioni, cioè a fenomeni in cui intervengano come variabili (indipendenti) due o più caratteri, (quantitativi o qualitativi).

Ora, la media aritmetica di n numeri reali è quel particolare valore medio A che risulta come parte aliquota n^a della loro somma o, più in generale, se i numeri sono dotati di certi pesi, è quel particolare valore medio che si ha quando la somma dei prodotti dei numeri pei rispettivi pesi venga divisa per la somma dei pesi (2). Si sa, d'altronde, che le sue proprietà caratteristiche sono: *a*) la somma algebrica dei prodotti degli scostamenti dei numeri dati dalla media aritmetica per i rispettivi pesi è nulla; *b*) la somma dei prodotti ottenuti dal moltiplicare per i rispettivi pesi i quadrati degli scostamenti dei numeri dati da un numero risulta minima quanto tale numero è la media aritmetica; *c*) la media aritmetica è, sopra una retta, il punto di applicazione della risultante di forze parallele applicate ai punti rappresentati dai numeri dati e aventi come intensità i pesi dati. E poichè queste proprietà sono indipendenti dalla esistenza di un minimo e di un massimo fra i numeri dati, così si potrà cercare la estensione del concetto di media aritmetica di una serie ciclica $Y = S\ ((X))$ in base a ciascuna di queste proprietà, di

(1) Cfr.: GINI. *The contributions of Italy to Modern Statistical Methods.* « Journal of the Royal Statistical Society », Vol. LXXXIX, Parte IV, 1926, pag. 718. Relazioni fra i diversi valori medî di più numeri dati possono determinarsi dipendentemente dalla considerazione di funzioni convesse, come fu osservato dallo JENSEN: *Sur les fonctions convexes et les inégalités entre les valeurs moyennes,* « Acta Mathematica », Tomo XXX; *v.* anche, a tale proposito, F. SIBIRANI, *Intorno alle funzioni convesse,* « Rend. R. Ist. Lomb, di Sc. e Lett. », 1907 e L. GALVANI, *Sulle funzioni convesse di una o due variabili definite in un aggregato qualunque,* in « Rend. del Cir. Mat. di Palermo », 1916.

(2) Affinchè la media ponderata di più numeri sia compresa fra il minimo ed il massimo di essi, è necessario che i pesi soddisfino a certe condizioni (*v.* in proposito C. G. BONFERRONI, *Sulle medie dedotte da funzioni concave,* in « Giorn. di Matem. Finanz. », 1927), ed è sufficiente che essi siano tutti positivi (Cfr. n. 8).

cui le prime due indicheremo brevemente con $\Sigma \varepsilon = o$, $\Sigma \varepsilon^2 = Min.$ Ammetteremo, salvo avviso contrario, che la Y non assuma mai valori negativi, ciò che certamente si verifica quando, p. es., la serie data sia una serie di frequenze.

Estensione del concetto di media aritmetica in base alla proprietà $\Sigma\varepsilon = o$

12. — Si è già detto che una serie ciclica qualunque $Y = S((X))$, si riduce praticamente ad una serie ciclica equispaziata, o per conveniente suddivisione del ciclo di variabilità, o mediante l'eventuale introduzione di modalità fittizie. Perciò, basterà considerare tra le serie cicliche quelle soltanto che siano equispaziate, intendendo che alle nuove modalità introdotte si facciano corrispondere valori nulli del carattere Y (e in particolare valori nulli della frequenza se si tratta di una serie di frequenze). Se a partire da una modalità X_1 che assumiamo convenzionalmente come prima e in uno dei due sensi, che diremo «positivo», le altre modalità della mutabile indipendente sono successivamente $X_2, X_3, \ldots X_s$, rappresenteremo graficamente tali modalità mediante punti omonimi equidistanti, $X_1, X_2, \ldots X_s$ di una circonferenza percorsa nel senso positivo (cioè in quello contrario al moto degli indici di un orologio). Alle modalità di X, ossia ai corrispondenti punti $X_1, X_2, \ldots X_s$ della circonferenza, dovranno intendersi rispettivamente associati i valori Y_1, $Y_2, \ldots Y_s$ della Y.

Alle «diversità base» fra le modalità $X_1, X_2, \ldots X_s$ di X corrisponderanno gli «scostamenti» dei punti (omonimi) X_1, X_2, $\ldots X_s$ che rappresentano quelle modalità sulla circonferenza; tali scostamenti saranno rappresentati da archi (non maggiori in valore assoluto di mezza circonferenza) aventi per estremi quei punti.

Riferendoci a tale rappresentazione potremo dire senz'altro «scostamenti» tra $X_1, X_2 \ldots X_s$ sia le diversità base fra le modalità, sia gli scostamenti fra i punti-modalità.

Per unità di diversità tra le modalità di X, e quindi come unità di scostamento tra i corrispondenti punti della circonferenza assumeremo, come già si disse, quella di due modalità consecutive, cosicchè condotto il diametro per X_1: se s è dispari, esso separerà gli $\frac{s-1}{2}$

punti che rispetto ad X_1 hanno scostamento positivo 1, 2, ... $\frac{s-1}{2}$, dagli $\frac{s-1}{2}$ a scostamento negativo —1, —2, ... — $\frac{s-1}{2}$; se s è pari, esso separerà ancora gli $\frac{s-2}{2}$ punti a scostamento positivo 1, 2, ... $\frac{s-2}{2}$ dagli $\frac{s-2}{2}$ a scostamento negativo —1, —2, ... — $\frac{s-2}{2}$, ed incontrerà il punto (opposto ad X_1) il cui scostamento è ad arbitrio $\pm\frac{s}{2}$.

Al cambiare della modalità che si assume come prima, muta la posizione di quel diametro, ma in qualunque posizione esso separa i punti a scostamento positivo da quelli a scostamento negativo, rispetto alla prima modalità.

Una modalità X_i *si dovrà dire la (od una) media aritmetica, nel senso fissato al principio di questo paragrafo, se la somma algebrica degli scostamenti delle altre modalità da* X_i, *rispettivamente moltiplicati per i corrispondenti valori di* Y, *è nulla (i valori di* Y *possono in particolare essere delle frequenze) cioè se*

$$\sum_{k=-\frac{s-1}{2}}^{\frac{s-1}{2}} k\,Y_{i+k} = Y_{i+1} + 2\,Y_{i+2} + \ldots + \frac{s-1}{2}\,Y_{i+\frac{s-1}{2}}$$

$$- Y_{i-1} - 2\,Y_{i-2} - \ldots - \frac{s-1}{2}\,Y_{i-\frac{s-1}{2}} =$$

$$= 0 \text{ (per } s \text{ dispari)}$$

o se

$$\sum_{k=-\frac{s-2}{2}}^{\frac{s-2}{2}} k\,Y_{i+k} = Y_{i+1} + 2\,Y_{i+2} + \ldots + \frac{s-2}{2}\,Y_{i+\frac{s-2}{2}}$$

$$- Y_{i-1} - 2\,Y_{i-2} - \ldots - \frac{s-2}{2}\,Y_{i-\frac{s-2}{2}} =$$

$$= 0 \text{ (per } s \text{ pari)}$$

convenendo, in quest'ultimo caso, che la modalità opposta a X_i, cioè la $X_{i} \mp \frac{s}{2}$ si scinda in due modalità, delle quali una di scostamento $\frac{s}{2}$ rispetto ad X_i ed avente $\frac{1}{2} Y_i + \frac{s}{2}$ come corrispondente

valore di Y, e l'altra di scostamento $-\frac{s}{2}$, e con un valore di Y dato pure da $\frac{1}{2} Y_i + \frac{s}{2}$ (1); e intendendo altresì che quando sia $p \equiv q \pmod{s}$ Y_p coincida con Y_q (Cfr. n. 10). Più brevemente si potrebbe dire che *una modalità* X_i *costituisce una media aritmetica se, rispetto ad essa, il momento primo della serie ciclica data,* $\mu_1(X_i)$, *si annulla.*

È facile dare esempî in cui tale definizione può essere verificata ed altri in cui essa rimane priva di significato.

Esempio 1°. — Si abbia una serie ciclica di frequenze, ottenuta distribuendo n eventi secondo gli $s = 7$ giorni della settimana: sia $n = 5$ eventi, dei quali 3 cadano nel *Me* e 2 nel *V*. Indicando con $\mu_1(L)$, $\mu_1(Ma)$... le somme algebriche degli scostamenti delle varie modalità rispettivamente dal lunedì, dal martedì, ecc., dopo avere attribuito come peso allo scostamento di ciascuna modalità il corrispondente valore di Y, si trova (2)

$$\mu_2(L) = 0 \; ; \; \mu_1(Ma) = 9 \; ; \; \mu_1(Me) = 4 \; ;$$
$$\mu_1(G) = -1 \; ; \; \mu_1(V) = -6 \; ; \; \mu_1(S) = -11 \; ; \; \mu_1(D) = 5.$$

La definizione assunta è dunque verificata per e soltanto per la modalità L, la quale si può perciò dire media aritmetica delle modalità del carattere qualitativo « giorno della settimana », nella considerata serie ciclica di frequenze.

(1) Cfr. C. Gini. *Variabilità e Mutabilità*, già cit.

(2) I reciproci scostamenti delle modalità della data serie sono messi in evidenza dal seguente prospetto:

da \ *a*	*L*	*Ma*	(1) *Me*	*G*	(2) *V*	*S*	*D*
L	0	1	2	3	—3	—2	—1
Ma	—1	0	1	2	3	—3	—2
Me	—2	—1	0	1	2	3	—3
G	—3	—2	—1	0	1	2	3
V	3	—3	—2	—1	0	1	2
S	2	3	—3	—2	—1	0	1
D	1	2	3	—3	—2	—1	0

Esempio 2°. — Sia $n = 4$ eventi, di cui 3 nel L e 1 nel Me; si trova

$$\mu_1(L) = 2 \; ; \; \mu_1(Ma) = -2 \; ; \; \mu_1(Me) = -6 \; ; \; \mu_1(G) = -10 \; ;$$
$$\mu_1(V) = 7 \; ; \; \mu_1(S) = 3 \; ; \; \mu_1(D) = 6.$$

La definizione non è verificata per nessuna modalità, cioè in questo caso non esiste media aritmetica nel senso detto.

Necessità dal punto di vista analitico di considerare il ciclo come continuo.

13. — Questi esempi dimostrano che, se ci si limita a considerare le sole modalità a scostamento intero del carattere X da cui dipende una serie ciclica, la definizione di media aritmetica che si è assunta può rimanere (e rimane generalmente) priva di significato.

Una completa trattazione analitica del problema esige dunque che, oltre alle modalità a scostamento intero altre se ne considerino, corrispondenti a scostamenti non interi, cioè esige che il ciclo di variabilità di X venga riguardato come continuo. Ciò consentirà non soltanto di ottenere, almeno formalmente, un risultato anche quando esso manca nel campo delle modalità effettivamente date, ma renderà altresì più facile l'applicazione dei procedimenti analitici. Circa l'interpretazione di un tale risultato converrà poi ricordare quanto si disse alla fine del n. 10, ma – insistiamo – la considerazione del ciclo di variabilità come continuo è in tutti i casi opportuna, anche indipendentemente dal significato reale che converrà attribuire al risultato. Perciò negli esempi che stiamo per dare ci limiteremo, in generale, alla ricerca dei risultati formali.

Diremo, per distinguere, modalità « effettive » di un carattere ciclico X quelle inizialmente date; modalità « concomitanti » quelle introdotte allo scopo di avere una serie equispaziata; modalità « di conto » quelle a scostamento non intero rispetto a una modalità effettiva o concomitante, ossia tutte le altre che si debbono pensare far parte di X, considerato come continuo.

Così nel primo esempio le modalità effettive sono Me e V, le concomitanti sono L, Ma, G, S, D. Ora, se consideriamo la modalità di conto P avente lo scostamento 4/5 da Me, troviamo

da \ a	L	Ma	$\overset{(3)}{Me}$	G	$\overset{(2)}{V}$	S	D
P	—14/5	—9/5	—4/5	1/5	6/5	11/5	16/5

e perciò $\mu_1(P) = 0$, cioè vediamo che alla definizione di media aritmetica ottempera, oltre alla modalità concomitante L, anche la modalità di conto P.

Nel secondo esempio, se Q è la modalità di conto avente da L lo scostamento $1/2$, risulta

da \ a	L	Ma	Me	G	V	S	D
Q	$-1/2$	$1/2$	$3/2$	$5/2$	$\pm 7/2$	$-5/2$	$-3/2$

onde $\mu_1(Q) = 0$ cioè la definizione di media aritmetica è verificata dalla modalità di conto Q, mentre non era verificata da nessuna modalità effettiva, nè concomitante.

Questi primi esempî mostrano intanto la possibilità che una serie ciclica ammetta parecchie medie aritmetiche, nel senso della definizione stabilita al principio, benchè in essi si prescinda, per ora, dai mezzi che si possono effettivamente impiegare per la ricerca delle medie stesse.

Medie ordinarie e medie «sui generis»

14. — Consideriamo un terzo esempio che darà luogo ad altri interessanti rilievi.

Esempio 3°. — Una serie ciclica di frequenze sia costituita da $n = 9$ eventi distribuiti come segue in $s = 8$ modalità:

Modalità di X:	S	SE	E	NE	N	NW	W	SW
Valori di Y:	0	0	3	0	4	2	0	0

Rispetto alla modalità di conto P avente lo scostamento $-7/9$ da SW, e alla modalità di conto Q di scostamento $5/9$ da NE si trova:

da \ a	S	SE	(3) E	NE	(4) N	(2) NW	W	SW
P	$16/9$	$25/9$	$34/9$	$-29/9$	$-20/9$	$-11/9$	$-2/9$	$7/9$
Q	$-32/9$	$-23/9$	$-14/9$	$-5/9$	$4/9$	$13/9$	$22/9$	$31/9$

da cui

$$\mu_1(P) = 0 \quad , \quad \mu_1(Q) = 0;$$

cioè la definizione di media è soddisfatta dalle due modalità di conto P e Q.

Ma vediamo ora quale sia la somma algebrica degli scostamenti delle modalità per cui non è nullo il valore di Y, cioè N, NW, E, dalle modalità rispettivamente opposte che indicheremo con N', $N'W'$, E', essendo $N' = S$, $N'W' = SE$, $E' = W$, e intendendo che ogni scostamento abbia come peso il rispettivo valore di Y. Si trova:

da \ a	(4) N	(2) NO	(3) E
$N' = S$	± 4	-3	2
$N'W' = SE$	3	± 4	1
$E' = O$	-2	-1	± 4

Se si ricorda la convenzione fatta (Cfr. n. 12), che nel caso di s pari la modalità opposta a quella presa come iniziale si debba intendere spezzata in due, aventi scostamenti opposti, e che a ciascuna di queste debba attribuirsi la metà del dato valore di Y, (ciò che equivale a supporre che il valore di Y attribuito alla modalità opposta a quella iniziale sia nulla) si trova:

$$\mu_1 (S) = 0 \ , \ \mu_1 (SE) = 15 \ , \ \mu_1 (O) = -10$$

cioè la modalità S ottempera alla posta definizione di media aritmetica.

Tuttavia esiste una differenza essenziale fra questa media S e le altre P e Q trovate in questo esempio ed anche quelle degli esempi precedenti).

Se si considera la somma algebrica $\mu_1 (X)$ degli scostamenti (opportunamente pesati) delle date modalità da una X fra le infinite modalità che il carattere X può assumere, considerandolo come una mutabile continua su tutto il ciclo, allora

$$\mu_1 (X)$$

sarà una funzione di X nel senso matematico della parola, qualora s'intenda, come abbiamo supposto, che le modalità di X siano rappresentate mediante punti omonimi sopra una circonferenza.

Ora è facile vedere che, nel nostro esempio, tale funzione è continua a destra e a sinistra dei punti immagini delle medie P e Q (e non ci indugiamo, per il momento, a mostrarlo), mentre ciò non si verifica per il punto immagine della media S.

Difatti, tendendo al punto $S = N'$ dalla sua sinistra (rispetto ad un osservatore che da S guardi il centro) lo scostamento limite

della modalità opposta N è -4, mentre tendendovi dalla destra lo scostamento limite di N è $+4$, e perciò

$$\lim_{X \ggg\to S} s \ \mu_1(X) = -4.Y_S = -16 \ ; \ \lim_{X \ggg\to S} d \ \mu_1(X) = +4.Y_S = 16$$

Il valore nullo della funzione $\mu_1(X)$ per $X = S$, in relazione alla convenzione fatta, *si presenta quindi* come la somma (ed è anzi preferibile dire *come la semisomma*) *dei valori a cui tende la stessa funzione a destra e a sinistra di* X.

Una media come S si dirà media aritmetica *sui generis*, per distinguerla da quelle già incontrate che si potranno dire « ordinarie ».

Medie limite

15. — Infine esaminiamo il seguente:

Esempio 4°. — Una serie ciclica di frequenze sia costituita da $n = 22$ eventi distribuiti nei giorni della settimana come appresso:

Modalità di X:	L	Ma	Me	G	V	S	D
Valori di Y:	5	2	1	6	3	0	5

Si trova facilmente che le modalità di conto P, Q, R, T, rispetto alle quali le date modalità hanno gli scostamenti indicati dal prospetto:

da \ a	(5) L	(2) Ma	(1) Me	(6) G	(3) V	S	(5) D
P	$-8/22$	$14/22$	$36/22$	$58/22$	$-74/22$	$-52/22$	$-30/22$
Q	$-29/22$	$-7/22$	$15/22$	$37/22$	$59/22$	$-73/22$	$-51/22$
R	$-64/22$	$-42/22$	$-20/22$	$2/22$	$24/22$	$-46/22$	$68/22$
T	$41/22$	$63/22$	$-69/22$	$-47/22$	$-25/22$	$-3/22$	$19/22$

verificano la definizione di media aritmetica perchè

$$\mu_1(P) = 0 \ , \ \mu_1(Q) = 0 \ , \ \mu_1(R) = 0 \ , \ \mu_1(T) = 0$$

Se poi consideriamo la modalità di conto $U = Ma'$ opposta a Ma, rispetto a cui gli scostamenti delle varie modalità sono

da \ a	L	Ma	Me	G	V	S	D
$Ma' = U$	$5/2$	$\pm 7/2$	$-5/2$	$-3/2$	$-1/2$	$1/2$	$3/2$

troviamo, in via del tutto analoga a quella seguita per la modalità S dell'esempio precedente: $\mu_1(U) = 7$, per la convenzione fatta di distribuire in parti uguali il valore $Y = 2$ fra la modalità Ma di scostamento $7/2$ e la modalità Ma di scostamento $= -7/2$; mentre i limiti a cui tende la $\mu_1(X)$ quando X tende ad U a sinistra e rispettivamente a destra sono:

$$\underset{X \ggg\!\!\rightarrow U}{\mathit{lims}}\ \mu_1(X) = 7 - \frac{14}{2} = 0\,, \quad \underset{X \ggg\!\!\rightarrow U}{\mathit{limd}}\ \mu_1(X) = 7 + \frac{14}{2} = 14.$$

Anche qui la funzione $\mu_1(X)$ è discontinua per $X = U$, ed il valore in tale punto è la semisomma dei limiti a cui la funzione stessa tende a sinistra e a destra di U. Uno di tali limiti, quello a sinistra è nullo; perciò si potrà dire che la modalità $Ma' = U$ costituisce una « media aritmetica limite a sinistra ». E similmente si potranno avere « medie aritmetiche limite a destra ».

Comportamento del momento primo $\mu_1(X)$

16. — Abbiamo voluto, di proposito, esporre anzitutto diversi esempi che ci dessero la possibilità di distinguere le varie specie di medie aritmetiche di una serie ciclica che, coerentemente alla definizione assunta, si possono presentare.

Una media aritmetica può dunque essere:

ordinaria: se la funzione $\mu_1(X)$ è in essa nulla e continua:	in senso stretto, se coincide con una delle date modalità di X; in senso lato, se non coincide con una di tali modalità;
sui generis:	se la $\mu_1(X)$ è in essa nulla e discontinua a destra e a sinistra;
limite a destra, o a sinistra:	se la $\mu_1(X)$ è in essa diversa da zero, ma tende allo zero a destra o, rispettivamente, a sinistra.

Si tratta ora di dare un procedimento atto alla determinazione di tali medie. Basterà, a tal fine, studiare il comportamento della funzione $\mu_1(X)$ al variare sulla circonferenza del punto immagine della modalità assunta dal carattere X, intendendo che tale

variazione abbia luogo con continuità, mentre le s modalità di X, effettive e concomitanti, costituiranno un insieme discreto equispaziato.

Diremo poi « momento di ordine zero » e indicheremo con μ_0 la somma $\sum_{i=1}^{s} Y_i$. In particolare, se la serie data fosse una serie di frequenze relative sarebbe $\mu_0 = \sum_{i=1}^{s} Y_i = 1$.

Siano ordinatamente nel senso positivo $X_1\, X_2 \ldots X_t\, X_{t+1} \ldots X_p$ i punti immagini di quelle $p \leq s$ modalità di X alle quali corrispondono valori non nulli di Y e tali valori siano $Y_1\, Y_2 \ldots Y_t\, Y_{t+1} \ldots Y_p$; mentre indicheremo con $X'_1\, X'_2 \ldots X'_t\, X'_{t+1} \ldots X'_p$ i punti rispettivamente opposti a quei primi. Naturalmente le modalità $X_1\, X_2 \ldots X_p$ non saranno, generalmente, equispaziate.

Supponiamo, intanto, s dispari, cosicchè $X'_1\, X'_2 \ldots X'_p$ saranno punti di mezzo degli archi determinati dalle coppie consecutive di quei punti che rappresentano le s modalità del ciclo.

Sia O una origine qualunque interna all'arco $X'_t\, X'_{t+1}$, X un punto qualunque, pure interno al medesimo arco, x ed x_i gli scostamenti di X e di X_i da O, essendo X_i una qualsiasi delle date modalità di X per cui non è nullo il corrispondente valore di Y; allora poichè i 3 punti O , X , X_i cadono tutti in uno stesso emiciclo determinato dal diametro $X_t\, X'_t$, o in uno stesso emiciclo determinato dal diametro X_{+1t} X'_{t+1} sarà, per una osservazione fatta alla fine del n. 9:

$$(XX_i) = (OX_i) - (OX) = x_i - x.$$

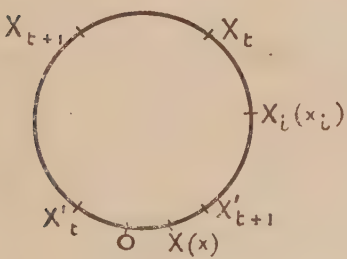

Fig. 1.

Perciò la funzione $\mu_1(X)$, di cui vogliamo constatare il comportamento, avrà in X il valore:

$$(1) \quad z(x) = \mu_1(X) = \sum_{i=1}^{p} (XX_i)\, Y_i = \sum_{i=1}^{p} x_i\, Y_i - $$
$$- x \sum_{i=1}^{p} Y_i = \mu_1(O) - \mu_0\, x$$

Essa è dunque una funzione lineare di x, rappresentata in un sistema cartesiano da una retta avente come ordinata all'origine O

il valore $\sum_{i=1}^{p} x_i Y_i = \mu_1(O)$ e come coefficiente angolare: $\sum_{i=1}^{p} Y_i = -\mu_o$, che si ridurrà alla unità negativa se la serie data è costituita da frequenze relative.

Se x è tale che risulti

$$z(x) = \sum_{i=1}^{p} x_i Y_i - x \sum_{i=1}^{p} Y_i = 0 \tag{1'}$$

cioè

$$x = \frac{\sum_{i=1}^{p} x_i Y_i}{\sum_{i=1}^{p} Y_i} = \frac{\mu_1(O)}{\mu_o}, \tag{2}$$

e se inoltre

$$x'_t < x < x'_{t=1} \tag{3}$$

essendo x'_t ed x'_{t+1} gli scostamenti di X'_t ed X'_{t+1} da O, allora il punto o modalità X corrispondente ad x costituirà una media aritmetica ordinaria della serie ciclica data, nel senso della posta definizione.

Se invece si verifica la (2) ed è:

$$\begin{matrix} x = x'_t \\ x = x'_{t+1} \end{matrix} \left\} \text{ la } X \text{ costituirà una media aritmetica limite } \right\{ \begin{matrix} \text{a destra} \\ \text{a sinistra} \end{matrix} \tag{4}$$

Se poi $x > x'_{t+1}$ oppure $x < x'_t$ non esiste nessuna media aritmetica del ciclo nell'intervallo da X'_t od X'_{t+1}.

Una conseguenza deriva subito da queste osservazioni, ed è che *le medie aritmetiche di una serie ciclica sono al più tante* quanti sono gli intervalli analoghi ad $X'_t X'_{t+1}$, cioè tante *quante sono le modalità del ciclo a cui corrispondono valori non nulli di* Y.

Altra conseguenza è che *entro ciascuno degli intervalli analoghi ad* $X'_t X'_{t+1}$ *la funzione* $\mu_1(X)$ *è* non soltanto lineare e quindi continua ma *rappresentata da un segmento di retta di coefficiente angolare essenzialmente negativo* ed uguale a $\sum_{i-1}^{p} Y_i$, qualunque sia l'intervallo.

Infine vediamo, riferendoci ad un sistema cartesiano nel quale l'asse delle ascisse porti le immagini delle modalità $X_1 X_2 \ldots X_p$ e delle modalità opposte $X'_1 X'_2 \ldots X'_p$, che se uno di quei segmenti incontra l'asse delle ascisse in un punto interno all'intervallo corrispondente $X'_t X'_{t+1}$, ivi si avrà la immagine di una media aritmetica ordi-

naria; se invece quel segmento incontra l'intervallo corrispondente $X'_t\ X_{t+1}$ nell'estremo sinistro o nell'estremo destro, in quel punto di incontro si avrà l'immagine di una media aritmetica limite a destra o rispettivamente a sinistra.

Rimane a vedere come si comporti la $\mu_1(X)$ nell'atto di attraversare uno degli estremi di $X'_t\ X'_{t+1}$, per esempio l'estremo X'_{t+1}.

Evidentemente, quando X è in X'_{t+1} la funzione $\mu_1(X)$ ha come valore la somma degli scostamenti delle varie modalità da X'_{t+1} moltiplicati pei rispettivi valori di Y, *tralasciando peraltro la modalità* X_{i+1} diametralmente opposta ad X'_{t+1} che, per una convenzione fatta, si deve considerare come scissa in due modalità aventi gli scostamenti di mezzo ciclo positivo e di mezzo ciclo negativo e aventi come corrispondenti valori di Y due quantità uguali a $\frac{1}{2} Y_{t+1}$.

Perciò

$$\mu_1(X'_{t+1}) = x_1 Y_1 + \ldots + x_t Y_t + x_{t+2} Y_{t+2} + \ldots + x_s Y_s,$$

intendendo, naturalmente, che gli scostamenti vengano misurati ancora dall'origine O.

Invece se X tende a X'_{t+1} dalla sua sinistra, la funzione $\mu_1(X)$ tende a quel valore che si ha aggiungendo a $\mu_1(X'_{t+1})$ il contributo della modalità X_{t+1} il cui scostamento da X'_{t+1} è al limite di mezzo ciclo negativo, cioè $-\frac{s}{2}$ e il cui corrispondente valore di Y è Y_{t+1}.

Perciò

$$\lim_{X \twoheadrightarrow X'_{t+1}} s\ \mu_1(X) = \mu_1(X'_{t+1}) - \frac{s}{2} Y_{t+1}. \tag{5}$$

Similmente si vedrebbe che

$$\lim_{X \twoheadrightarrow X'_{t+1}} d\ \mu_1(X) = \mu_1(X'_{t+1}) + \frac{s}{2} Y_{t+1}. \tag{6}$$

Possiamo dunque concludere che la funzione $\mu_1(X)$, nell'atto di attraversare il punto X'_{t+1}, salta dal valore (5) al valore (6), offrendo una discontinuità misurata da

$$\lim_{X \twoheadrightarrow X'_{t+1}} d\ \mu_1(X) - \lim_{X \twoheadrightarrow X'_{t+1}} s\ (X) = sY_{t+1} \tag{7}$$

mentre il valore di $\mu_1(X)$ in X'_{t+1} è la semisomma dei due limiti (5) e (6) a sinistra e a destra, cioè

$$(8) \qquad \mu_1(X'_{t+1}) = \frac{1}{2}\left(\underset{X \ggg\!\!\to X'_{t+1}}{lims}\ \mu_1(X) + \underset{X \ggg\!\!\to X'_{t+1}}{limd}\ \mu_1(X)\right).$$

Tutte le considerazioni sin qui fatte possono ripetersi per il caso che il numero s delle modalità di X sia pari, con la sola modificazione che i punti $X'_1\, X'_2 \ldots X'_p$, diametralmente opposti ai punti immagini delle modalità per le quali non è nullo il corrispondente valore di Y, coincidono rispettivamente coi punti immagini di altre modalità.

Osserviamo anche, dalla (8), che, comunque sia la parità di s, quando

$$(9) \qquad \mu_1(X'_{t+1}) = o\,,$$

il che richiede che si abbia

$$\underset{X \ggg\!\!\to X'_{t+1}}{lims}\ \mu_1(X) = -\underset{X \ggg\!\!\to X'_{t+1}}{limd}\ \mu_1(X)$$

il punto X'_{t+1} sarà immagine di una media aritmetica *sui generis*. Se poi s è pari, una tale media *sui generis* coinciderà con una delle date modalità della serie ciclica proposta.

Notiamo altresì che in tutte le relazioni testè scritte le sommatorie si intendono estese a quei soli p valori dell'indice i a cui corrispondono valori non nulli di Y, ma esse conserverebbero gli stessi valori anche estese alla totalità degli s valori dell'indice.

Infine osserviamo che se la radice x dell'equazione 1') non soddisfa alla limitazione

$$x'_t \leq x \leq x'_{t+1}$$

essa non rappresenta, come abbiamo detto, nè una media aritmetica ordinaria, nè limite, ma la sua immagine sarebbe data dalla intersezione con l'asse delle ascisse del *prolungamento* di quel segmento che in corrispondenza all'intervallo da X'_t ad X'_{t+1} rappresenta lo andamento della funzione $\mu_1(X)$; per generalità di linguaggio, diremo che in tal caso x rappresenta una media aritmetica «fittizia».

Riassumendo, in corrispondenza a ciascun intervallo come $X'_t \ldots X'_{t+1}$ si avrà o una media ordinaria o una media limite, o una media fittizia. Le medie fittizie non sono, peraltro, modalità di X per cui si annulli $\mu_1(X)$, ma conviene tenerle presenti per quanto sarà detto più oltre (*Vedi* n. 18).

Oltre a tali medie si potranno poi avere delle medie *sui generis* nei punti di separazione fra due segmenti consecutivi come $X'_t \ldots X'_{t+1}$

Designeremo complessivamente le medie ordinarie e le medie *sui generis* come medie « accettabili » perchè in esse il momento primo si annulla pure essendo rispettivamente continuo e discontinuo; le medie limite e le medie fittizie si diranno, invece, « inaccettabili » perchè in esse il momento primo non è zero (pure tendendo da una parte a zero nelle medie limite).

I diversi tipi di medie aritmetiche si presentano graficamente come segue:

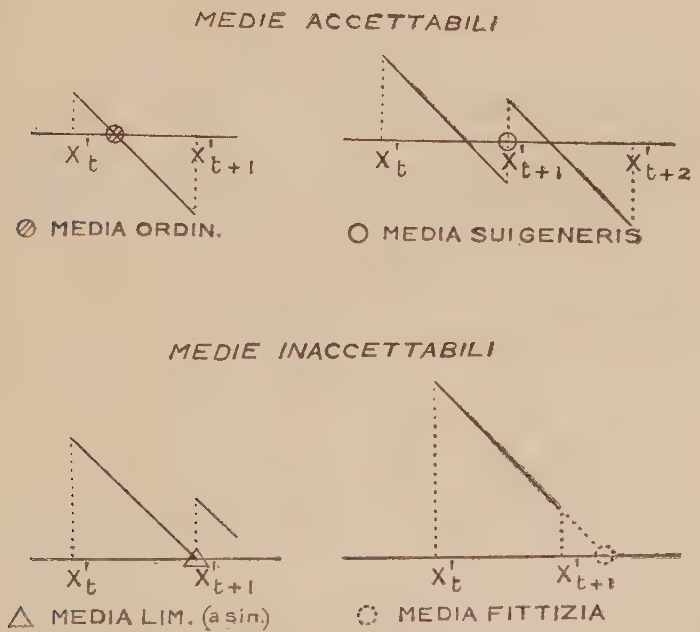

Fig. 2.

Il grafico annesso all'esempio 8 (*Vedi* n. 19) dà le immagini di medie ordinarie in P, Q, R, T, U; di una media limite a sinistra in N; di una media fittizia in F; di una media *sui generis* in E.

17. — Passando dall'intervallo $X'_t\ X'_{t+1}$ al successivo $X'_{t+1}\ X'_{t+2}$ sarebbe necessario assumere una nuova origine $\bar{O}$ interna a questo, e allora, essendo X un punto dello stesso intervallo, si avrebbe, come prima

$$z = \mu_1(X) = \sum_{i=1}^{p} \bar{x}_i\, Y_i - \bar{x} \sum_{i=1}^{p} Y_i$$

intendendo che $\bar{x}$ ed $\bar{x}_i$ siano gli scostamenti di X e di X_i dalla origine $\bar{O}$.

E se poi si avesse

$$z = 0\,, \text{ cioè } \bar{x} = \frac{\sum_{i=1}^{p} \bar{x}_i\, Y_i}{\sum_{i=1}^{p} Y_i}$$

con la condizione

$$\bar{x}'_{t+1} \leq \bar{x} \leq \bar{x}'_{t+2}$$

allora $\bar{x}$ rappresenterebbe una nuova media aritmetica, non esterna all'intervallo $X'_{t+1}\ X'_{t+2}$ e quindi ordinaria o limite, mentre se tale limitazione non si verificasse la $\bar{x}$, rappresenterebbe una media fittizia. Ma la determinazione di una tale media si può eseguire anche più semplicemente notando che *se si conserva la stessa origine* O *utilizzata nell'intervallo* $X'_t\ X'_{t+1}$ allora, poichè la $\mu_1(X)$ ha andamento lineare e presenta nel punto X'_{t+1} che separa i due intervalli la discontinuità sY_{t+1} mantenendo la direzione costante di coefficiente angolare $-\sum_{i=1}^{p} Y_i$, così entro il secondo intervallo la $\mu_1(X)$ sarà rappresentata da

$$(10)\quad z = \Sigma x_i\, Y_i - x\, \Sigma Y_i + sY_{t+1} = \mu_1(O) - x\, \mu_0 + sY_{t+1}$$

Passando all'intervallo $X'_{t+2}\ X'_{t+3}$ e conservando ancora l'origine O si avrà similmente

$$z = \mu_1(O) - x\,\mu_0 + s\,Y_{t+1} + s\,Y_{t+2}',$$

e così via, finchè pervenendo all'intervallo $X'_{t-1}\ X'_t$ che immediatamente precede il primo considerato $X'_t\ X_{t+1}$ si troverà

$$z = \mu_1(O) - x\,\mu_0 + s(Y_{t+1} + Y_{t+2} + \ldots + Y_{t-1})\,.$$

Oltrepassando poi X_t si troverà

$$z = \mu_1(O) - x\mu_o + s\sum_{i=1}^{p} Y_i \,. \tag{11}$$

Naturalmente in tutte queste formule, dalla (10) alla (11), bisognerà ricordare che $X_t = X_r$ e similmente $Y_t = Y_r$ se $t \equiv r$ (mod. s); ed anche che, dopo aver percorso l'intero ciclo di mutabilità di X, lo scostamento x, *che si è sempre continuato a misurare da* O, si sarà accresciuto di s unità rispetto al primitivo cosicchè, sostituendo ad x il valore primitivo, la (11) si potrà scrivere

$$\begin{aligned} z &= \mu_1(O) - x\mu_o - s\mu_o + s\mu_o \\ &= \mu_1(O) - x\mu_o \end{aligned} \tag{11'}$$

che coincide con la (1).

Ciò posto, per la ricerca della media corrispondente all'intervallo $X'_{t+1}\,X'_{t+2}$, utilizzando la (10), cioè riferendosi sempre all'origine O, si avrà:

$$x = \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i} + \frac{s\, Y_{t+1}}{\Sigma\, Y_i} = \frac{\mu_1(O)}{\mu_o} + \frac{s\, Y_{t+1}}{\mu_o}$$

la quale sarà o non sarà accettabile secondo che la limitazione

$$x'_{t+1} \leq \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i} + \frac{s\, Y_{t+1}}{\Sigma\, Y_i} \leq x'_{t+2} \tag{12}$$

sia oppure non sia verificata.

Similmente la media corrispondente all'intervallo $X'_{t+2}\,X'_{t+3}$ sarà data, conservando l'origine O, da

$$x = \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i} + \frac{s\, Y_{t+1}}{\Sigma\, Y_i} + \frac{s\, Y_{t+2}}{\Sigma\, Y_i}\,,$$

e così via, fintantochè la media corrispondente all'intervallo iniziale sarà data da

$$x = \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i} + \frac{s\,\Sigma\, Y_i}{\Sigma\, Y_i} = \frac{\mu_1(O)}{\mu_o} + s$$

che non differisce, naturalmente, da quella che avevamo trovato essere rappresentata da

$$x = \frac{\mu_1(O)}{\mu_o}\,.$$

18. — I valori

$$(13) \qquad \frac{s\,Y_{t+1}}{\Sigma\,Y_i}, \frac{s\,Y_{t+2}}{\Sigma\,Y_i}, \ldots, \frac{s\,Y_{t-1}}{\Sigma\,Y_i}, \frac{s\,Y_t}{\Sigma\,Y_i}$$

sono dunque tali che addizionati successivamente a $\frac{\Sigma\,x_i\,Y_i}{\Sigma\,Y_i}$ (che rappresentava lo scostamento da O della media corrispondente allo intervallo iniziale $X'_t\,X'_{t+1}$) forniscono via via gli scostamenti da O delle medie (ordinarie, limite o fittizie) corrispondenti agli intervalli analoghi successivi, fino a ritornare alla media iniziale.

Viene così confermato che le medie aritmetiche ordinarie sono al più tante quanti sono quegli intervalli, ossia quante sono le modalità del ciclo per cui non è nullo il corrispondente valore di Y. Naturalmente, oltre alle medie ordinarie, possono aversi come medie accettabili anche quelle *sui generis* le quali, se esistono, cadono nei punti $X'_1\,X'_2\ldots X'_p$ opposti ai punti $X_1\,X_2\ldots X_p$ pei quali Y non è nullo.

Ma è soprattutto da rilevare il fatto che *lo scostamento da* O *di ciascuna media ordinaria, limite o fittizia si ottiene dallo scostamento della media che compete all'intervallo precedente addizionando a questo ultimo scostamento il prodotto di* $\frac{s}{\Sigma\,Y_i}$ *per il valore di* Y *corrispondente alla modalità* X_{t+1} *di cui si deve attraversare il punto opposto* X'_{t+1} *per andare dall'intervallo relativo alla prima media a quello relativo alla seconda.*

Va anche notato che in questo procedimento di passaggio da una media alla successiva, si debbono via via considerare non soltanto quelle ordinarie corrispondenti, come si è detto, alle intersezioni dell'asse delle ascisse con i segmenti che rappresentano la $\mu_1\,(X)$ entro i successivi intervalli determinati dai punti $X'_1\,X'_2\ldots X'_p$ e quelle limiti a sinistra o a destra, ma anche quelle fittizie i cui scostamenti da O non verificano le limitazioni analoghe alla 12) e che, come abbiamo detto, sono geometricamente rappresentate dalle intersezioni dei prolungamenti di quei segmenti con l'asse delle ascisse. Al contrario, il *procedimento descritto non utilizza e non fornisce senz'altro le medie* sui generis.

In conclusione, l'andamento complessivo della $\mu_1\,(X)$ è rappresentato geometricamente in modo assai semplice quando gli scostamenti delle modalità, misurati con continuità da una origine fissa O, si assumano come ascisse di punti di una retta.

Nei diversi intervalli come $X'_t\, X'_{t+1}$ la $\mu_1(X)$ è rappresentata da segmenti di retta aventi tutti la stessa direzione, data dal coefficiente angolare $-\Sigma\, Y_i = -\mu_0$ e che tagliano o non tagliano l'asse delle ascisse secondo che nell'intervallo cada o non cada una media aritmetica ordinaria. La discontinuità in ciascun estremo fra un segmento e quello successivo è data dal prodotto $s\, Y_{t+1}$ del numero s delle modalità per il valore di Y corrispondente alla modalità opposta a quell'estremo. Quei segmenti che non tagliano l'asse delle ascisse ma che hanno comune con tale asse un estremo, forniscono, con questo estremo, l'immagine di una media limite a sinistra o a destra. Se l'estremo destro di un segmento e quello sinistro del successivo sono equidistanti dall'asse delle ascisse si ha ivi una media aritmetica *sui generis*. Se un segmento non incontra l'asse delle ascisse, il suo prolungamento incontra tale asse in un punto che non rappresenta una media aritmetica accettabile, ma una media fittizia. Le medie ordinarie e limite e quelle fittizie sono successivamente separate da intervalli misurati dai valori del sistema (13).

Osservazione 1ª. — Una media aritmetica nel senso ristretto considerato in principio (Cfr. n. 12), cioè tale da coincidere con una delle modalità del ciclo inizialmente date, può avere o il carattere di una media aritmetica ordinaria, o quello di una media aritmetica *sui generis*. Nel primo caso essa non deve essere opposta ad un'altra modalità del ciclo, cioè deve essere s dispari; oppure, se s è pari, deve opporsi ad una modalità cui compete un valore nullo di Y. Nell'altro caso essa deve necessariamente opporsi ad una modalità per cui sia diverso da o il valore di Y e quindi deve essere s pari.

Osservazione 2ª. — Una media *sui generis* è caratterizzata dal fatto che in essa i limiti sinistro e destro di $\mu_1(X)$ sono opposti (e quindi il limite a sinistra è essenzialmente negativo e quello a destra essenzialmente positivo). Perciò, avendo la $\mu_1(X)$ decorso lineare con coefficiente angolare $-\mu_0$, le due medie, ordinarie, limite o fittizie limitrofe a quella media *sui generis*, sono da questa equidistanti; e reciprocamente. Ne viene che, quando gli scostamenti delle medie siano tutti misurati da una stessa origine e in uno stesso senso, come ai nn. 17 e 18, *le medie* sui generis *sono determinate da quegli scostamenti che sono la semisomma degli scostamenti di due medie, effettive o fittizie, corrispondenti a segmenti consecutivi, e che denotano uno dei punti* $X'_1\, X'_2 \ldots X'_p$. *Una volta trovate, con qualunque pro-*

cedimento, le medie ordinarie, limite e fittizie, sarà dunque facile determinare anche le medie sui generis.

Osservazione 3ª. — Poichè le medie aritmetiche ordinarie, limite e fittizie sono successivamente separate da intervalli le cui ampiezze sono date dal sistema (13), così si può dire che le medie stesse sono rappresentate da punti che dividono la circonferenza in parti direttamente proporzionali ai p valori non nulli di Y; precisamente se $X_1 X_2 \ldots X_p$ sono le modalità di X cui corrispondono valori non nulli di Y e cioè $Y_1 Y_2 \ldots Y_p$, ed $X'_1 X'_2 \ldots X'_p$ sono i punti diametralmente opposti a quei primi, determinata la media aritmetica (ordinaria, o limite, o fittizia) corrispondente ad un intervallo qualunque, p. es., $X'_1 X'_2$, tutte le altre si otterranno nel senso positivo come punti di divisione della circonferenza in parti proporzionali ai numeri $Y_2 Y_3 \ldots Y_{p-1} Y_p Y_1$ a partire da quella prima media.

Se ne deduce che, nell'ipotesi di s pari, se le modalità per cui è $Y \neq 0$ costituiscono una configurazione simmetrica e se inoltre tutti i valori di Y sono uguali, i punti di mezzo delle coppie di modalità consecutive costituiscono medie aritmetiche ordinarie della serie ciclica data, mentre i punti modalità costituiscono delle medie *sui generis*. Se invece s è dispari e si conservano le altre ipotesi, i punti modalità costituiscono delle medie ordinarie, e i punti di mezzo delle coppie di modalità consecutive rappresentano delle medie *sui generis*.

Se rispetto ad un punto della circonferenza (e quindi anche rispetto all'opposto) i punti modalità e i corrispondenti valori di Y sono disposti simmetricamente, quel punto rappresenta una media aritmetica ordinaria se il punto opposto non sia immagine di una modalità per cui Y è nullo; e rappresenta, invece, una media aritmetica *sui generis* se il punto opposto è immagine di una modalità per cui Y non sia nullo.

Infine due serie cicliche, aventi le stesse modalità e valori di Y proporzionali nelle modalità corrispondenti, hanno le stesse medie aritmetiche ordinarie, limite, *sui generis* e fittizie.

Osservazione 4ª. — Se tutte le modalità $X_1 X_2 \ldots X_t$ alle quali corrispondono valori non nulli di Y giacciono in uno stesso emiciclo, gli intervalli a ciascuno dei quali competerà una media ordinaria o limite o fittizia, saranno al più in numero di t e saranno determinati dai punti $X'_1 X'_2 \ldots X'_t$. Fra tali intervalli quello $X'_t \ldots X'_1$, *percorso nello stesso ordine delle* $X_1 X_2 \ldots X_t$, sarà almeno uguale ad un emiciclo, e conterrà tutte queste modalità. La media aritmetica

relativa all'intervallo stesso $X'_t \ldots X'_1$ avrà, rispetto ad una origine O, comunque scelta internamente ad esso, uno scostamento:

$$\bar{x} = \frac{\mu_1(O)}{\mu_o} = \frac{\sum_{i=1}^{t} x_i \, Y_i}{\sum_{i=1}^{t} Y_i}$$

e poichè $Y_i > O$ $(i = 1, \ldots t)$ risulterà

$$x_1 < \bar{x} < x_t \, .$$

Quindi: in $X'_t \ldots X'_1$ cade una sola media $\overline{X}$; essa si determina, rispetto ad una origine ivi scelta, nello stesso modo che per le seriazioni e per le serie rettilinee; ed essendo compresa fra la prima e l'ultima delle $X_1 \ldots X_t$ (come nel caso delle seriazioni e delle serie rettilinee per le quali sia $Y_i > O$) costituisce una media ordinaria.

Naturalmente altre medie, accettabili o no, corrisponderanno agli altri intervalli $X'_1 \ldots X'_2$, ...; $X'_{t-1} \ldots X'_t$; ma *la sola media assimilabile a quella che si avrebbe per una seriazione o per una serie rettilinea, quanto all'essere compresa fra la prima e l'ultima delle modalità $X_1 \ldots X_t$ è quella $\overline{X}$ ora segnalata,* della quale vedremo più oltre una importante proprietà.

Due metodi di ricerca delle medie aritmetiche

19. — Le considerazioni dei numeri precedenti ci consentono di dare due regole molto semplici per la determinazione delle medie aritmetiche di una serie ciclica.

Regola 1ª. — Si può, anzitutto, dividere la circonferenza rappresentativa del ciclo di mutabilità della X in p intervalli mediante i punti $X'_1\, X'_2 \ldots X'_p$ rispettivamente opposti alle modalità $X_1\, X_2 \ldots X_p$ per cui siano $(Y_1, Y_2, \ldots Y_p) \neq o$, fissare in ciascuno di questi intervalli una origine arbitraria O, determinare gli scostamenti delle varie modalità rispetto a tale origine e applicare la formula

$$x = \frac{\Sigma \, x_i \, Y_i}{\Sigma \, Y_i} = \frac{\mu_1(O)}{\mu_o}$$

la quale rappresenterà la media aritmetica ordinaria, limite o fittizia corrispondente a tale intervallo. E precisamente: una media

ordinaria se, rispetto alla stessa origine, x costituisce lo scostamento di un punto interno al medesimo intervallo; una media limite se x è lo scostamento di un suo estremo; una media fittizia se x è lo scostamento di un punto ad esso esterno. Se l'origine è il punto di mezzo dell'intervallo, secondo che x sia in valore assoluto $\lesseqgtr \frac{1}{2}$, la media rappresentata sarà ordinaria, limite o fittizia. Calcolate in tal modo le medie che competono ai successivi intervalli, si provi se esistono anche medie *sui generis:* basta, a tal fine, calcolare i valori di $\mu_1(X)$ in corrispondenza a ciascuno dei punti X'_i nella supposizione che il valore Y_i corrispondente ad X_i sia nullo, e le medie *sui generis*, se esistono, saranno quelle per cui $\mu_1(X)$ risulta nullo; esse potranno anche trovarsi come è indicato dalla osservazione 2ª del n. 18.

Esponiamo due applicazioni di questo metodo, una per s pari, ed una per s dispari.

Esempio 5°. — Si riprenda la serie ciclica dell'esempio 4°, e cioè:

Modalità di X:	L	Ma	Me	G	V	S	D
Valori di Y:	5	2	1	6	3	0	5

Gli intervalli da considerare sono i sei determinati dai punti

$$L' \quad Ma' \quad Me' \quad G' \quad V' \quad D' \quad (L')$$

entro ciascuno dei quali potremo prendere come origine, per semplicità, dato che s è dispari, il rispettivo punto modalità, e cioè ordinatamente:

$$V \quad S \quad D \quad L \quad Ma \quad G$$

Osservando lo specchio degli scostamenti

da \ a	(5) L	(2) Ma	(1) Me	(6) G	(3) V	(5) D
V	3	—3	—2	—1	0	2
S	2	3	—3	—2	—1	1
D	1	2	3	—3	—2	0
L	0	1	2	3	—3	—1
Ma	—1	0	1	2	3	—2
G	—3	—2	—1	0	1	3

si ha subito:

da $V: x_{\mathrm{I}} = \dfrac{15 - 6 - 2 - 6 + 10}{22} = \dfrac{11}{22}$, corrispondente ad una media limite a sinistra, perchè tale media cade all'estremo destro dell'intervallo L' Ma';

da $S: x_{\mathrm{II}} = \dfrac{10 + 6 - 3 - 12 - 3 + 5}{22} = \dfrac{3}{22}$, scostamento di una media ordinaria T, perchè interna all'intervallo che si considera Ma' Me';

da $D: x_{\mathrm{III}} = -\dfrac{12}{22}$, non accettabile, perchè $\dfrac{12}{22}$ è lo scostamento di un punto esterno all'intervallo Me' G' (media fittizia);

da $L: x_{\mathrm{IV}} = \dfrac{8}{22}$, scostamento di una media ordinaria P;

da $Ma: x_{\mathrm{V}} = \dfrac{7}{22}$, scostamento di una media ordinaria Q;

da $G: x_{\mathrm{VI}} - = \dfrac{2}{22}$, scostamento di una media ordinaria R.

La serie data ammette dunque quattro medie aritmetiche ordinarie ed una media limite a sinistra.

Si trova poi che

$$\mu_1 (L') = 4{,}5\,, \ \mu_1 (Ma') = 7\,, \ \mu_1 (Me') = -4{,}5\,,$$
$$\mu_1 (G') = -2\,, \ \mu_1 (V') = 7{,}5\,, \ \mu_1 (D') = -8{,}5\,.$$

Poichè nessuno di questi valori è nullo, la data serie non ammette nessuna media aritmetica *sui generis*.

Esempio 6°. — Una serie ciclica di «intensità» sia costituita da $n = 8$ eventi distribuiti secondo $s = 8$ modalità cicliche, per ciascuna delle quali l'evento considerato abbia una certa intensità Y:

Modalità di X:	S	SE	E	NE	N	NW	W	SW
Valore di Y (intensità):	3	5	0	4	2	7	3	6

Si vede che le medie aritmetiche saranno 7 al più, e che gli intervalli da considerare saranno determinati da S', $S'E'$, $N'E'$, N' $N'W'$, W', $S'W'$, S'; per semplicità, entro ciascuno di questi prenderemo come origine il punto di mezzo fra due consecutive delle date modalità. Perciò questi punti indicheremo con

$$SSE, \quad ESE, \quad ENE, \quad NNE, \quad NNW, \quad WNW, \quad SSW$$

e rispetto ad essi gli scostamenti delle varie modalità con valori non nulli di Y saranno:

da \ a	(3) S	(5) SE	(4) NE	(2) N	(7) NW	(3) W	(6) SW
SSE	— 0,5	0,5	2,5	3,5	— 3,5	— 2,5	— 1,5
ESE	— 1,5	— 0,5	1,5	2,5	3,5	— 3,5	— 2,5
ENE	— 2,5	— 1,5	0,5	1,5	2,5	3,5	— 3,5
NNE	— 3,5	— 2,5	— 0,5	0,5	1,5	2,5	3,5
NNW	3,5	— 3,5	— 1,5	— 0,5	0,5	1,5	2,5
WNW	2,5	3,5	— 2,5	— 1,5	— 0,5	0,5	1,5
SSW	0,5	1,5	3,5	— 3,5	— 2,5	— 1,5	— 0,5

cosicchè risulterà:

$$\text{da } SSE\colon x_{\mathrm{I}} = \frac{-3+5+20+14-49-15-18}{2 \times 30} = -\frac{46}{60},$$

non accettabile;

da ESE: $x_{\mathrm{II}} = \frac{6}{60}$, scostamento di una media ordinaria;

da ENE: $x_{\mathrm{III}} = -\frac{6}{60}$, scostamento di una media ordinaria;

da NNE: $x_{\mathrm{IV}} = \frac{30}{60}$, scostamento di una media limite a sinistra;

da NNW: $x_{\mathrm{V}} = \frac{18}{60}$, scostamento di una media ordinaria;

da WNW: $x_{\mathrm{VI}} = \frac{38}{60}$, scostamento di una media ordinaria;

da SSW: $x_{\mathrm{VII}} = -\frac{18}{60}$, scostamento di una media ordinaria.

Esistono dunque cinque medie aritmetiche ordinarie ed una media limite a sinistra.

Si trova poi

$$\mu_1(S') = 12,\ \mu_1(S'E') = 14,\ \mu_1(N'E') = -10,\ \mu_1(N') = -16$$
$$\mu_1(N'W') = -10,\ \mu_1(W') = 0,\ \mu_1(S'W) = 6$$

e quindi si conclude che W' costituisce una media *sui generis*.

Regola 2ª. — Determinata, col metodo già esposto, la media aritmetica ordinaria, o limite, o fittizia, relativa ad uno dei soliti intervalli, p. es. ad $X'_t\, X'_{t+1}$, tutte le altre medie, comprese quelle limite e quelle fittizie, ma escluse quelle *sui generis*, si otterranno

dalla prima, nell'ordine positivo *e con riferimento alla medesima origine già scelta*, mediante successive addizioni dei valori

$$\frac{s\, Y_{t+1}}{\Sigma\, Y_i}\,,\quad \frac{s\, Y_{t+2}}{\Sigma\, Y_i}\,,\ldots,\frac{s\, Y_{t-1}}{\Sigma\, Y_i}$$

i cui numeratori costituiscono, come già sappiamo, i salti della funzione $\mu_1(X)$ nei suoi punti di discontinuità, e cioè in X'_{t+1}, X'_{t+2}, ecc.

Come riprova della esattezza dei calcoli, l'aggiunta ulteriore della discontinuità $s\, Y_t$ dovrà fornire di nuovo la media aritmetica di partenza.

Infine gli scostamenti delle medie *sui generis* saranno quelle semisomme degli scostamenti di due medie, ordinarie, limite o fittizie consecutive, le quali corrispondono a uno dei punti $X'_1\; X'_2 \ldots X'_p$; e quindi se nessuna di tali semisomme verifica questa condizione, non esisterà nessuna media *sui generis*.

Osservazione. — Sarà facile dare, con riferimento ad un sistema cartesiano ortogonale sul cui asse delle ascisse siano rappresentate le modalità di X, una semplice rappresentazione grafica del momento primo:

$$z = \mu_1(X)\,,$$

la quale rappresentazione risulterà ulteriormente semplificata assumendo come unità di misura non già lo scostamento fra due modalità consecutive del ciclo, ma la sua parte aliquota $\dfrac{1}{\sum\limits_{i=1}^{s} Y_i}$, poichè allora, come risulta dalla (1') del n. 16, il coefficiente angolare dei segmenti che rappresentano con le ordinate dei loro punti la funzione $\mu_1(X)$ sarà —1, cioè tali segmenti saranno inclinati di —45° rispetto all'asse delle ascisse.

Illustriamo ora anche la 2ª regola con alcuni esempi:

Esempio 7°. — Data la serie ciclica:

Modalità di X:	L	Ma	Me	G	V	S	D
Valori di Y:	5	2	0	7	0	0	0

e considerati gli intervalli determinati da L', Ma', G', (L'), si trovi la media corrispondente a uno di essi, p. es., a $L'\,Ma'$, assumendo in questo come origine V; lo scostamento di tale media sarà allora:

$$\text{da } V\!:\; x = \frac{2}{14} \qquad \text{(media ordinaria in } L'\,Ma'\text{)}.$$

Procedendo in senso positivo, le medie successive avranno dalla stessa origine degli scostamenti che si otterranno da quello trovato $x = \frac{2}{14}$ per successive addizioni dei valori $\frac{s\,Y_i}{\Sigma\,Y_i}$, i quali sono proporzionali ai successivi salti di $\mu_1\,(X)$. Perciò, dato che nel nostro caso $\frac{s}{\Sigma\,Y_i} = \frac{7}{14}$, tutte le medie della serie ciclica data risulteranno dal seguente prospetto:

INTERVALLO	MODALITÀ opposta a quella che si attraversa	VALORE di Y in tale modalità	$\frac{s\,Y_i}{\Sigma\,Y_i}$	SCOSTAMENTO della media corrispondente da V
$L' \ldots Ma'$	..	..	..	$x_{\mathrm{I}} = \frac{2}{14}$
$Ma' \ldots G'$	Ma	2	$\frac{7}{14} \cdot 2 = \frac{14}{14}$	$x_{\mathrm{II}} = \frac{2}{14} + \frac{14}{14} = \frac{16}{14} = 1 + \frac{2}{14}$
$G' \ldots L'$	G	7	$\frac{7}{14} \cdot 7 = \frac{49}{14}$	$x_{\mathrm{III}} = \frac{16}{14} + \frac{49}{14} = \frac{65}{14} = 4 + \frac{9}{14}$
PROVA: $L'Ma'$	L	5	$\frac{7}{14} \cdot 5 = \frac{35}{14}$	$x_{\mathrm{I}} = \frac{65}{14} + \frac{35}{14} = 7 + \frac{2}{14}$

Gli scostamenti trovati definiscono 3 medie ordinarie Q, R, P.

Lo scostamento $7 + \frac{2}{14}$ definisce la media di partenza Q. Fatte le semisomme delle successive coppie di valori

$$\frac{2}{14}\,,\; 1 + \frac{2}{14}\,,\; 4 + \frac{9}{14}\,,\; \left(7 + \frac{2}{14}\right)$$

le quali risultano essere

$$\frac{9}{14}\,,\; 2 + \frac{12{,}5}{14}\,,\; 5 + \frac{12{,}5}{14}$$

si constata che nessuna di queste rappresenta lo scostamento, rispetto a V, di uno dei punti $L'\ Ma'\ G'$; perciò la data serie non ammette medie aritmetiche *sui generis*.

Infine, la rappresentazione grafica di $\mu_{\mathrm{I}}(X)$ è la seguente:

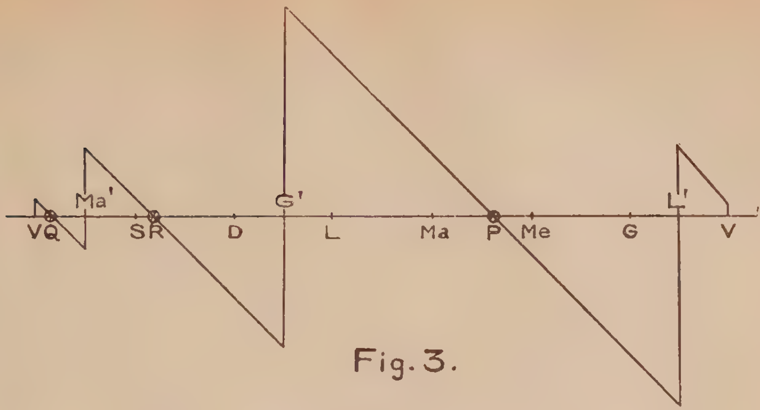

Fig. 3.

Si noti che le modalità a cui corrispondono valori non nulli di Y, e cioè L, Ma, G giacciono in uno stesso emiciclo; la media (ordinaria) compresa fra L e G (indicata con P e avente da V lo scostamento x_{III}) è quella particolare media a cui si riferisce l'osservazione 4ª del n. 18.

Esempio 8º. — Si riprenda la serie ciclica dall'esempio 6º, in cui si è trovato che la media competente all'intervallo $N' \ldots N'W'$ (cioè all'intervallo $S \ldots SE$) è determinata dallo scostamento:

$$\text{da } SSE\colon x = -\frac{23}{30}$$

(media fittizia relativa a $N' \ldots N'W'$).

Tutte le medie successive incontrate nel senso positivo, saranno perciò date dal seguente prospetto, per la cui costruzione bisogna tenere presente che $\dfrac{s}{\Sigma Y_i} = \dfrac{8}{30}$ e che l'origine è costantemente SSE:

Intervallo	Modalità opposta a quella che si attraversa	Valore di Y in tale modalità	$\frac{s\,Y_i}{\Sigma\,Y_i}$	Scostamento da $S\,S\,E$ della media corrispondente all'intervallo.
$N' \dots N'W'$	..	..	..	$x_{\mathrm{I}} = -\frac{23}{30}^{(1)}$
$N'W' \dots W'$	$N\,W$	7	$\frac{56}{30}$	$x_{\mathrm{II}} = -\frac{23}{30} + \frac{56}{30} = \frac{33}{30} = 1 + \frac{3}{30}^{(2)}$
$W' \dots S'W'$	W	3	$\frac{24}{30}$	$x_{\mathrm{III}} = \frac{33}{30} + \frac{24}{30} = \frac{57}{30} = 1 + \frac{27}{30}^{(2)}$
$S'' \dots SW'$	$S\,W$	6	$\frac{48}{30}$	$x_{\mathrm{IV}} = \frac{57}{30} + \frac{48}{30} = \frac{105}{30} = 3 + \frac{15}{30}^{(3)}$
$S' \dots S'E'$	S	3	$\frac{24}{30}$	$x_{\mathrm{V}} = \frac{105}{30} + \frac{24}{30} = \frac{129}{30} = 4 + \frac{9}{30}^{(2)}$
$S'E' \dots N'E'$	$S\,E$	5	$\frac{40}{30}$	$x_{\mathrm{VI}} = \frac{129}{30} + \frac{40}{30} = \frac{169}{30} = 5 + \frac{19}{30}^{(2)}$
$N'E' \dots N'$	$N\,E$	4	$\frac{32}{30}$	$x_{\mathrm{VII}} = \frac{169}{30} + \frac{32}{30} = \frac{201}{30} = 6 + \frac{21}{30}^{(2)}$
Prova: $N' \dots N'O'$	N	2	$\frac{16}{30}$	$x_{\mathrm{I}} = \frac{201}{30} + \frac{16}{30} = \frac{217}{30} = 8 - \frac{23}{30}$

(1) Scostamento di una media fittizia F (non accettabile).
(2) Scostamenti di medie ordinarie P, Q, R, T, U.
(3) Scostamento di una media limite a sinistra S'.

In E si ha una media *sui generis* (che è anche una media in senso stretto), perchè E è punto di discontinuità della $\mu_1(X)$ corrispondente alla semisomma degli scostamenti delle medie limitrofe: $\frac{1}{2}\left(\frac{33}{30} + \frac{57}{30}\right) = 1{,}5$ (scostamento di E dalla solita origine $S\,SE$).

La rappresentazione grafica di μ_1 (X) è la seguente:

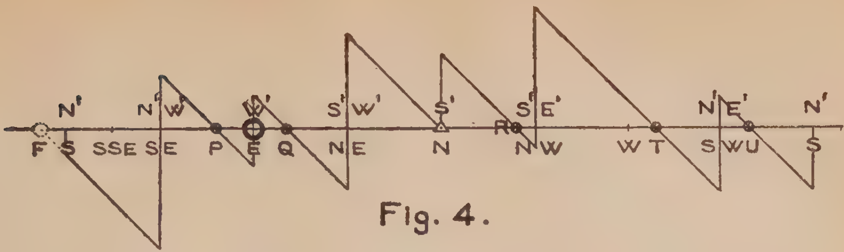

Fig. 4.

Interpretazione della numerosità delle medie aritmetiche

20. — Il fatto che una serie ciclica ad s modalità, che supponiamo sempre equispaziate, può ammettere una media aritmetica ordinaria in corrispondenza ad ognuno degli s intervalli in cui il ciclo risulta scomposto dai punti opposti ai punti modalità, e quindi al più s medie aritmetiche ordinarie, ci induce a domandarci quale sia il significato da attribuire alla maggiore o minore numerosità delle medie ordinarie che una data serie ciclica può ammettere.

Ora, poichè una media ordinaria è tale che in essa il momento primo della serie si annulla, cioè le modalità dell'emiciclo precedente e quelle dell'emiciclo seguente alla media considerata forniscono rispetto alla media stessa momenti uguali in valore assoluto, così essa costituisce, *in tal senso*, un centro di simmetria della serie data. D'altronde abbiamo veduto (Cfr. n. 18 osserv. 3ª) che distribuzioni cicliche simmetriche nel senso che alle diverse modalità di X corrispondano uguali valori di Y, forniscono una media ordinaria in corrispondenza a (e nel punto di mezzo di) ciascuno degli intervalli sopra detti. Perciò il fatto che di due serie cicliche date in funzione dello stesso numero s di modalità di uno stesso carattere X, una ammette un maggior numero di medie aritmetiche ordinarie che non l'altra, può essere interpretato come un indizio che la prima serie ciclica è « più simmetrica » dell'altra. In base a tale concetto si potrebbe anche dare una misura della simmetria di una serie ciclica; e su questo argomento ci ripromettiamo di ritornare in altra occasione.

ESTENSIONE DEL CONCETTO DI MEDIA ARITMETICA IN BASE ALLA PROPRIETÀ $\Sigma\varepsilon^2 = \text{MIN.}$

21. — Considereremo, anche in questo caso, le sole serie cicliche equispaziate, rese eventualmente tali mediante l'introduzione di modalità concomitanti, e ci riferiremo alla solita rappresentazione grafica delle s modalità $X_1\ X_2 \ldots X_s$ mediante punti di una circonferenza, intendendo che a tali modalità corrispondano rispettivamente i valori $Y_1\ Y_2 \ldots Y_s$ della variabile Y.

Una modalità X_i *si dovrà dire media aritmetica nel senso che rispetto ad essa la somma dei quadrati degli scostamenti risulti minima, se tale risulti la somma dei quadrati degli scostamenti delle altre modalità da* X_i *rispettivamente moltiplicati per i corrispondenti valori di* Y, cioè se:

$$\sum_{k=-\frac{s-1}{2}}^{\frac{s-1}{2}} k^2\, Y_{i+k} = (Y_{i+1} + Y_{i-1}) + 2^2\,(Y_{i+2} + Y_{i-2}) + \ldots + \left(\frac{s-1}{2}\right)^2 \left(Y_{i+\frac{s-1}{2}} + Y_{i-\frac{s-1}{2}}\right) = \text{Min. (per } s \text{ dispari)}$$

$$\text{e} \quad \sum_{k=-\frac{s}{2}}^{\frac{s}{2}} k^2\, Y_{i+k} = (Y_{i+1} + Y_{i-1}) + 2^2(Y_{i+2} + Y_{i-2}) + \ldots + \left(\frac{s}{2} - 1\right)^2 \left(Y_{i+\frac{s-2}{2}} + Y_{i-\frac{s-2}{2}}\right) + \left(\frac{s}{2}\right)^2 \left(\frac{1}{2} Y_{i+\frac{s}{2}} + \frac{1}{2} Y_{i+\frac{s}{2}}\right) = \text{Min. (per } s \text{ pari)}$$

intendendosi, come sempre, in questo ultimo caso, che la modalità opposta ad X_i, cioè la $X_{i\pm\frac{s}{2}}$ si scinda in due modalità delle quali una di scostamento $\frac{s}{2}$ e l'altra di scostamento $-\frac{s}{2}$ rispetto ad X_i e aventi, l'una e l'altra, $\frac{1}{2} Y_{i+\frac{s}{2}}$ come corrispondente valore di Y; e intendendo altresì che se $p \equiv q$ (mod. s), sia $Y_p = Y_q$.

A differenza della definizione di media aritmetica basata sullo annullamento della somma algebrica degli scostamenti (Cfr. n. 12) la quale, quando si voleva riferire limitatamente ad una effettiva modalità X_i del ciclo poteva perdere ed anzi generalmente perdeva significato, *la definizione ora posta non perde mai significato, perchè,*

determinata la somma dei prodotti di Y pei quadrati degli scostamenti delle altre modalità rispetto a ciascuna modalità del ciclo, *esisterà sempre una di queste, almeno, per cui quella somma costituisce il minimo assoluto*, senza escludere che la somma stessa abbia valore costante (se $Y_1 = Y_2 = \ldots = Y_s$), nel qual caso ciascuna modalità soddisferà alla condizione di minimo, cioè costituirà una media aritmetica nel senso testè dichiarato.

Esempio 9°. — Si riprenda la serie ciclica di frequenze costituita da $n = 22$ eventi distribuiti tra i giorni della settimana secondo lo specchio:

Modalità di X:	L	Ma	Me	G	V	S	D	$(s = 7)$
Modalità di Y:	5	2	1	6	3	0	5	$(n = 22)$

Formato il solito specchio dei reciproci scostamenti delle varie modalità è facile calcolare la $\Sigma k^2 Y_{i+k}$ rispetto a ciascuna modalità e si trova

$$\sum_{k=-3}^{3} k^2 Y_{L+k} = (2 + 5) + 2^2 (1 + 0) + 3^2 (6 + 3) = 92 ,$$

$$\Sigma k^2 Y_{Ma+k} = 77 , \quad \Sigma k^2 Y_{Me+k} = 85 , \quad \Sigma k^2 Y_{G+k} = 102 ,$$

$$\Sigma k^2 Y_{V+k} = 93 , \quad \Sigma k^2 Y_{S+k} = 79 , \quad \Sigma k^2 Y_{D+k} = 88 .$$

Poichè fra queste somme la minima 77 è quella ottenuta prendendo come origine degli scostamenti la modalità Ma si dovrà dire che, nel senso della posta definizione, la media aritmetica è costituita dalla modaltà Ma.

Si constata dunque che, in questo senso, la serie ciclica proposta ammette una media aritmetica; mentre, secondo la prima definizione la serie stessa non ammetteva nessuna media aritmetica in senso stretto, cioè coincidente con una modalità effettiva (Cfr. n. 13), e ne ammetteva quattro in senso lato, cioè quando si assumeva che una media aritmetica potesse anche essere costituita da una modalità di conto (non coincidente con nessuna delle date modalità).

Viene dunque naturale di ricercare se anche la seconda definizione di media aritmetica possa assumere un significato più generale, non riferendola soltanto alle modalità effettive, ma anche alle modalità di conto del ciclo.

Perciò anche qui, a somiglianza di quanto si è fatto partendo dalla prima definizione di media aritmetica, si potrà considerare il sistema delle modalità del carattere qualitativo X come parte di un ciclo continuo, di cui diremo ancora « modalità di conto » quelle a

scostamento non intero rispetto a una modalità effettiva o concomitante (Cfr. n. 13); ed allora vi sarà luogo a considerare per ogni modalità generica X del ciclo continuo la somma dei quadrati degli scostamenti delle modalità $X_1\, X_2 \ldots X_s$ da X, rispettivamente moltiplicate per $Y_1\, Y_2 \ldots Y_s$, cosicchè risulterà definita per tutto il ciclo, e quindi per ogni punto della circonferenza rappresentativa del carattere X, la funzione:

$$\zeta = \mu_2\,(X) = \sum_{i=1}^{s} (X X_i)^2\, Y_i\,,$$

essendo $X X_i$ lo scostamento di X_i da X.

Ci proponiamo di studiare il comportamento di questa funzione, che potremo dire «momento secondo» della serie data rispetto ad X, e che vedremo essere una funzione continua di X, ciò che ci consentirà di mettere in relazione i suoi minimi, assoluti e relativi, con la nuova definizione di media aritmetica che abbiamo assunto.

Il significato di tale definizione risulterà così ampliato nel senso che: i minimi relativi della $\mu_2\,(X)$ si verificheranno in corrispondenza ad una o a più modalità effettive o no del ciclo, e tali modalità costituiranno altrettante medie aritmetiche, a termini della seconda definizione ampliata; volendo, si potranno poi particolarmente considerare fra tutte queste medie quella o quelle sole in cui $\mu_2\,(X)$ raggiunge il suo minimo assoluto, e qualora una di queste medie coincida con una modalità effettiva, essa sarà anche una media nel senso della seconda definizione non ampliata. Le considerazioni che seguono si riferiscono, salvo avviso contrario, alla seconda definizione in senso lato, cioè contemplano tutte quelle modalità di X in cui $\mu_2\,(X)$ sia un minimo relativo.

Notiamo poi che, secondo la prima definizione, potevano esistere medie aritmetiche *sui generis* coincidenti coi punti opposti alle modalità per cui non fosse nullo il corrispondente valore di Y, e risultanti dalla circostanza che rispetto a tali punti presi come origini le modalità opposte davano luogo allo scostamento semplice:

$$\frac{s}{2}\,\frac{1}{2}\,Y_{i+\frac{s}{2}} - \frac{s}{2}\,\frac{1}{2}\,Y_{i+\frac{s}{2}} = 0\,;$$

mentre nel senso della seconda definizione non possono esistere medie *sui generis* perchè quegli stessi punti danno luogo allo scostamento quadratico

$$\left(\frac{s}{2}\right)^2 \left(\frac{1}{2}\,Y_{i+\frac{s}{2}} + \frac{1}{2}\,Y_{i+\frac{s}{2}}\right) \neq 0\,.$$

Per maggiore chiarezza, eseguiamo nel seguente prospetto un parallelo fra le due definizioni di media aritmetica.

I DEFINIZIONE $\mu_1(X) = 0$	in senso stretto	Riferibile alle sole modalità effettive del ciclo; generalmente non esiste nessuna di tali medie.	
	in senso lato	Riferibile a qualunque modalità del ciclo, effettiva o no.	medie ordinarie. medie non accettabili cioè m. limiti e m. fittizie.
	Medie *sui generis*.		
II DEFINIZIONE $\mu_2(X) = \text{Min.}$	in senso stretto	Riferibile alle sole modalità effettive del ciclo; esiste sempre almeno una media.	
	in senso lato	Riferibile a qualunque modalità del ciclo, effettiva o no.	*a*) medie corrispondenti ai minimi relativi della $\mu_2(X)$. *b*) medie corrispondenti al minimo o ai minimi assoluti della $\mu_2(X)$.
	Non esistono medie *sui generis*.		

Comportamento del momento secondo $\mu_2(X)$.

22. — Assumendo ancora le stesse ipotesi e notazioni già impiegate al n. 16 e senza ormai più distinguere il caso di s dispari da quello di s pari, se O ed X sono una origine arbitraria ed una modalità generica non esterne all'intervallo $X'_t\, X'_{t+1}$ si avrà come valore della funzione $\mu_2(X)$ in X:

$$(1)\qquad \zeta(x) = \mu_2(X) = \sum_{i=1}^{p} (X X_i)^2\, Y_i = \sum_{i=1}^{p} (x_i - x)^2\, Y_i =$$

$$= x^2 \sum_{1}^{p} Y_i - 2x \sum_{1}^{p} x_i\, Y_i + \sum_{1}^{p} x_i^2\, Y_i = \mu_0\, x^2 - 2\,\mu_1(O)\,.\,x + \mu_2(O)\ (1)\,.$$

(1) Come al n. 16, se la data fosse una serie di frequenze relative, sarebbe $\Sigma\, Y_i = \mu_0 = 1$.

Riferendosi a un sistema cartesiano ortogonale la (1) rappresenta dunque internamente ad $X'_t\ X'_{t+1}$ un arco di parabola disposto superiormente all'asse delle ascisse e volgente quindi verso lo stesso asse la sua convessità.

Affinchè X sia punto di minimo della (1) è necessario e sufficiente che si abbia

$$\zeta'(x) = 2\,\mu_0\,x - 2\,\mu_1\,(O) = 0\,,\ \zeta''(x) = 2\,\mu_0 > 0$$

cioè che sia

$$x = \frac{\mu_1\,(O)}{\mu_0}\,,\ \mu_0 > 0 \tag{2}$$

la quale disuguaglianza si verifica per ipotesi, e che inoltre

$$x'_t < x < x'_{t+1} \tag{3}$$

Le condizioni (2) e (3) sono quelle stesse che determinano una media aritmetica ordinaria, nel senso della prima definizione, internamente all'intervallo $X'_t\ X'_{t+1}$ cosicchè concludiamo, intanto, che *le medie corrispondenti ai minimi relativi, nel senso lato della seconda definizione, sono le medie aritmetiche ordinarie a norma della prima definizione.*

Se

$$x = x'_t \quad \text{oppure} \quad x = x'_{t+1}\,, \tag{4}$$

il detto arco di parabola ha un minimo nell'uno o nell'altro estremo di $X'_t\ X'_{t+1}$ in corrispondenza a quella che, secondo la prima definizione, era una media limite a destra o a sinistra.

Infine, se nè la (3) nè la (4) sono verificate, quell'arco di parabola, cioè $\mu_2\,(X)$ non ha un minimo *internamente* ad $X'_t\ X'_{t+1}$.

Supponiamo, ora, di attraversare il punto X'_{t+1}. L'espressione del momento secondo, cioè la (1) risulterà modificata in ciò, che mentre X si avvicina ad X'_{t+1}, lo attraversa e lo oltrepassa (in un intorno convenientemente ristretto) lo scostamento di X_{t+1} da X crescerà assolutamente per valori negativi fino a $-\frac{s}{2}$, salterà da $-\frac{s}{2}$ ad $\frac{s}{2}$, e poi decrescerà da $\frac{s}{2}$, cosicchè il quadrato di tale scostamento, che entra nella espressione di $\mu_2\,(X)$, si manterrà continuo; lo stesso $\mu_2\,(X)$ si manterrà anche continuo in quella sua parte che dipende da tutte le altre modalità $X_1\ X_2 \ldots X_t\ X_{t+2} \ldots X_p$: perciò concludiamo che *nell'atto di attraversare* X'_{t+1}, *cioè il punto opposto ad una*

modalità, il momento secondo $\mu_2(X)$ *conserva un decorso continuo, anche se in quella modalità il corrispondente valore di* Y *sia diverso da zero.*

Inoltre, osservando che lo scostamento da X ad X_{t+1} è:

$$(X\,X_{t+1}) = x_{t+1} - x$$

fintanto che X sia nell'emiciclo precedente X'_{t+1} nel senso positivo, mentre diviene:

$$(X\,X_{t+1}) = x_{t+1} - x + s$$

quando X passi nell'emiciclo seguente ad X'_{t+1}, si concluderà che, sostituendo nella espressione di $\mu_2(X)$ valida in $X'_t \ldots X'_{t+1}$ alla quantità:

$$(x_{t+1} - x)^2\, Y_{t+1} = (x^2_{t+1} - 2x\, x_{t+1} + x^2)\, Y_{t+1}$$

la quantità

$$(x_{t+1} - x + s)^2\, Y_{t+1} = (x^2_{t+1} - 2x\, x_{t+1} + x^2 + 2s\, x_{t+1} - \\ - 2s\, x + s^2)\, Y_{t+1}$$

si otterrà senz'altro l'espressione di $\mu_2(X)$ valida nell'intervallo successivo a quel primo, cioè in $X'_{t+1} \ldots X'_{t+2}$ conservando sempre il riferimento alla origine O.

Pertanto tale espressione di $\mu_2(X)$ sarà

$$(5) \quad \zeta(x) = \mu_2(X) = \mu_0\, x^2 - 2\,(\mu_1(O) + s\, Y_{t+1})x + \\ + (\mu_2(O) + s^2\, Y_{t+1} + 2s\, x_{t+1}\, Y_{t+1}),$$

equazione, anche questa, di un arco di parabola che si connette al primo in corrispondenza ad X'_{t+1}.

Similmente dalla (5) si passerà all'espressione di $\mu_2(X)$ valida nell'intervallo $X'_{t+2} \ldots X'_{t+3}$ mediante l'aggiunta di

$$-2s\, Y_{t+2} \quad \text{e di} \quad s^2\, Y_{t+2} + 2s\, x_{t+2}\, Y_{t+2}$$

rispettivamente al coefficiente di x e al termine noto; e così via, fino a pervenire all'espressione di $\mu_2(X)$ in $X'_{t-1} \ldots X'_t$ che si avrebbe da quella che la precede in $X'_{t-2} \ldots X'_{t-1}$ con le rispettive addizioni di

$$-2s\, Y_{t-1} \quad \text{e di} \quad s^2\, Y_{t-1} + 2s\, x_{t-1}\, Y_{t-1}$$

al secondo e all'ultimo coefficiente. Così si potrà conoscere l'espressione di $\mu_2(X)$ per qualsiasi modalità X del ciclo.

Eseguendo una volta di più l'operazione indicata, e cioè aggiungendo al coefficiente di x e al termine noto di $\mu_2(X)$ rispettivamente:

$$-2s\,Y_t \quad \text{e} \quad s^2\,Y_t + 2s\,x_t\,Y_t$$

si otterrà

$$\mu_0\,x^2 - 2\,(\mu_1(O) + s\,\mu_0)\,x + (\mu_2(O) + s^2\,\mu_0 + 2s\,\mu_1(O)) =$$
$$= \mu_0\,(x-s)^2 - 2\mu_1(O)\,(x-s) + \mu_2(O)\,,$$

che è l'espressione di $\mu_2(X)$ già ottenuta per l'intervallo iniziale $X'_t \ldots X'_{t+1}$, poichè, ovviamente, avendo conservato l'origine O, lo scostamento di X si è accresciuto di s dopo un giro.

Concludendo, la funzione $\mu_2(X)$ è geometricamente rappresentata in un sistema cartesiano ortogonale da un insieme di archi di parabola che si stendono, superiormente all'asse della x, in corrispondenza ai successivi intervalli come $X'_t \ldots X'_{t+1}$ e che si connettono, ciascuno al seguente, negli estremi di tali intervalli, senza, però, raccordarsi fra di loro, cioè senza avere, negli estremi stessi, la tangente comune.

Difatti, derivando rispetto ad x la (1) e la (4) si ha rispettivamente:

$$\zeta'(x) = 2\,\mu_0\,x - 2\,\mu_1(O)$$
$$\zeta'(x) = 2\,\mu_0\,x - 2\,\mu_1(O) - 2s\,Y_{t+1}$$

le quali dimostrano che nell'atto di attraversare X'_{t+1} in senso positivo la $\zeta'(x)$ presenta una discontinuità misurata da $-2\,s\,Y_{t+1}$.

Per maggiore chiarezza, si potrà subito osservare l'andamento del momento secondo di una serie ciclica nella Fig. 5 relativa all'esempio 10° del numero seguente.

Notiamo poi che la conoscenza di $\mu_2(X)$ lungo l'intero ciclo permette di determinare rapidamente lo scostamento quadratico medio della serie rispetto a ciascuna delle sue medie aritmetiche, (ed anche, eventualmente, rispetto ad altre modalità). Difatti lo scostamento quadratico medio della serie rispetto ad una modalità qualsiasi $\overline{X}$ è evidentemente:

$$^2\eta_{\overline{X}} = \left(\frac{\mu_2(\overline{X})}{\Sigma\,Y_i}\right)^{\frac{1}{2}}$$

Ricerca delle medie

23. — Siamo ora in grado di esaurire il problema di ricerca delle modalità del ciclo rispetto alle quali risulti minimo il momento secondo, cioè di ricerca delle medie aritmetiche nel senso della seconda definizione.

Se vogliamo limitare la nostra indagine alle medie in senso stretto, cioè alle sole modalità effettive ottemperanti alla condizione di minimo richiesta, è ovvio che basterà calcolare il momento secondo rispetto a ciascuna delle s modalità, e constatare per quale o per quali modalità tale momento acquista il suo minimo valore. Esiste sempre almeno una media in questo senso; ma essa non è, di regola, una media a tenore della prima definizione, neanche nel suo senso stretto.

Invece, le medie aritmetiche in senso lato, coincidenti o no con una modalità effettiva, potranno definirsi sia in corrispondenza ai minimi relativi, sia in corrispondenza al minimo assoluto di $\mu_2(X)$. Come sappiamo già, le prime fra tali medie, sono quelle che corrispondono agli eventuali punti di minimo degli archi di parabola già considerati *e coincidono con le medie aritmetiche ordinarie nel senso della prima definizione.* Volendo poi fra tali medie scegliere quella o quelle corrispondenti al minimo dei minimi, basterà calcolare il momento secondo della data serie rispetto a ciascuna di esse, e constatare a quali modalità competa il minimo assoluto. Un mezzo per calcolare i momenti secondi rispetto a ciascuna media (ordinaria nel senso della prima definizione) può consistere appunto nel determinare le equazioni degli archi di parabola corrispondenti ai successivi intervalli come $X'_t \ldots X'_{t+1}$, e nel valutare le ordinate di tali archi nei punti indici di quelle medie.

Potrebbe sorgere il dubbio che il punto comune a due di quegli archi di parabola, fra loro consecutivi, potesse costituire il punto di ordinata minima per l'uno e per l'altro arco, e in tal caso il punto stesso (estremo di un intervallo come $X'_t \ldots X'_{t+1}$) corrisponderebbe ad un minimo relativo di $\mu_2(X)$ che non sarebbe fornito dall'analisi precedente; ma basta riflettere che in quel punto la derivata $\zeta'(x)$ dovrebbe, come negli altri suoi punti di discontinuità, fare un salto del valore $-2\,s\,Y_{t+1}$ cioè essenzialmente negativo, per concludere che oltrepassando il punto X_{t+1} la funzione $\mu_2(X)$ dovrebbe continuare a decrescere e che perciò quella eventualità non può cer-

tamente presentarsi; cosicchè il problema di ricerca delle medie nel senso della seconda definizione, non ammette altre soluzioni all'infuori di quelle segnalate.

Vogliamo anche osservare che in corrispondenza a quelle che erano, nel senso della prima definizione, medie limite a sinistra o a destra, si hanno dei punti in cui uno dei due archi di parabola ivi limitati possiede la tangente parallela all'asse delle ascisse.

Infine, notiamo anche che le medie *sui generis*, nel senso della prima definizione, vengono a corrispondere a dei punti nei quali si congiungono due di quei soliti archi di parabola simmetrici fra loro rispetto ad un asse passante per quel punto, normalmente all'asse delle ascisse. Infatti nell'intorno a sinistra e a destra di quel punto, il momento primo, ossia, all'infuori di un fattore, la derivata del momento secondo assume valori contrarî, come sappiamo dal n. 18, (osserv. 2ª).

La seconda parte del prospetto dato al n. 21 si potrà dunque completare come segue:

<table>
<tr>
<td rowspan="4">II
DEFINIZIONE
$\mu_2(X)=$Min.</td>
<td>In senso stretto</td>
<td colspan="2">Riferibile alle sole modalità effettive del ciclo; esiste sempre almeno una di tali medie, che non è generalmente una media a tenore della I definizione</td>
</tr>
<tr>
<td rowspan="2">In senso lato</td>
<td rowspan="2">Riferibile a qualunque modalità del ciclo, effettiva o di conto</td>
<td>*a*) Medie corrispondenti ai minimi relativi della $\mu_2(X)$: sono tutte le medie ordin. a termini della I definizione</td>
</tr>
<tr>
<td>*b*) Medie corrispondenti al minimo o ai minimi assoluti della $\mu_2(X)$: sono comprese fra le precedenti e quindi comprese fra le medie ordinarie a termini della I definizione</td>
</tr>
<tr>
<td colspan="3">Non esistono medie *sui generis*; e le medie *sui generis* a termini della I definizione, non sono medie a termini della II definizione</td>
</tr>
</table>

Esempio 10°. — Consideriamo di nuovo la serie ciclica degli esempi 4°, 5° e 9°, e cioè:

Modalità di X:	L	Ma	Me	G	V	S	D
Valori di Y:	5	2	1	6	3	0	5

Si è già veduto (es. 9°) che l'unica media nel senso stretto della seconda definizione è costituita dalla modalità Ma per cui risulta $\mu_2\ (Ma) = 77$.

A norma della stessa definizione, le medie in senso lato corrispondenti ai minimi relativi di $\mu_2(X)$ sono le 4 medie ordinarie nel senso della prima definizione, già determinate nell'esempio 5°.

Perciò occupiamoci ora delle medie in senso lato corrispondenti al minimo assoluto di $\mu_2(X)$; potremo, a tal fine, calcolare i momenti secondi relativi a quelle 4 medie e prescegliere quella o quelle per cui tale momento è minimo. Ma preferiamo considerare le equazioni dei successivi archi di parabola, e scegliere la o le modalità di X a cui corrisponde il minimo assoluto delle ordinate, ciò che ci consentirà anche di dare la rappresentazione geometrica di $\mu_2(X)$.

Assumendo nell'intervallo G' V' l'origine L, l'equazione del soprastante arco di parabola sarà $\zeta(x) = \mu_0 x^2 - 2\mu_1(L)\,x + \mu_2(L)$, i cui coefficienti si potranno così calcolare:

		(5)	(2)	(1)	(6)	(3)	(5)	
	da \ a	L	Ma	Me	G	V	D	$s = 7$ $\mu_0 = 22$
Scostam:	L	0	1	2	3	-3	-1	$\mu_1(L) = 8$
Quad. degli scost.	L	0	1	4	9	9	1	$\mu_2(L) = 92$

Pertanto, ricordando il procedimento di passaggio dalla equazione di un arco di parabola a quella dell'arco di parabola successiva, si potrà costruire il seguente prospetto:

Intervallo	**Modalità** opposta a quella che si attraversa	**Valore** di Y in tale modalità	$-2s\,Y_i$ da aggiungere al 2° coefficiente	$s^2\,Y_i + 2s\,x_i\,Y_i$ da aggiungere al 3° coefficiente	**Equazione** dell'arco di parabola	**Scostamenti** delle medie aritmetiche ordinarie da L (v. es. 5°)	**Momento** secondo in tali medie
$G' \ldots V'$	..	..	..	..	$\zeta = 22x^2 - 16x + 92$	$x_{\text{I}} = \frac{8}{22}$	$\zeta\left(\frac{8}{22}\right) = \frac{1960}{22}$
$V' \ldots D'$	V	3	$-14.3 = -42$	$49.3 + 14(-3)\,3 = 21$	$\zeta = 22x^2 - 58x + 113$	$x_{\text{II}} = 1 + \frac{7}{22} = \frac{29}{22}$	$\zeta\left(\frac{29}{22}\right) = \frac{1645}{22}$
$D' \ldots L'$	D	5	-70	$49.5 + 14(-1)\,5 = 175$	$\zeta = 22x^2 - 128x + 288$	$x_{\text{III}} = 3 - \frac{2}{22} = \frac{64}{22}$	$\zeta\left(\frac{64}{22}\right) = \frac{2240}{22}$
$L' \ldots Ma'$	L	5	-70	$49.5 + 14.0.5 = 245$	$\zeta = 22x^2 - 198x + 533$	..	..
$Ma' \ldots Me'$	Ma	2	-28	$49.2 + 14.1.2 = 126$	$\zeta = 22x^2 - 226x + 659$	$x_{\text{V}} = 5 + \frac{3}{22} = \frac{113}{22}$	$\zeta\left(\frac{113}{22}\right) = \frac{1729}{22}$
$Me' \ldots G'$	Me	1	-14	$49.1 + 14.2.1 = 77$	$\zeta = 22x^2 - 240x + 736$	..	..
Prova $G' \ldots V'$	G	6	-84	$49.6 + 14.3.6 = 546$	$\zeta = 22x^2 - 324x + 1282$	..	..

Osservando l'ultima colonna di questo prospetto, si vede che la media aritmetica nel senso lato della seconda definizione, corrispondente al minimo assoluto del momento secondo, è unica e coincidente con quella media aritmetica nel senso della prima definizione il cui scostamento da L è $\frac{29}{22}$.

Diamo infine la rappresentazione grafica di $\mu_2\,(X)$.

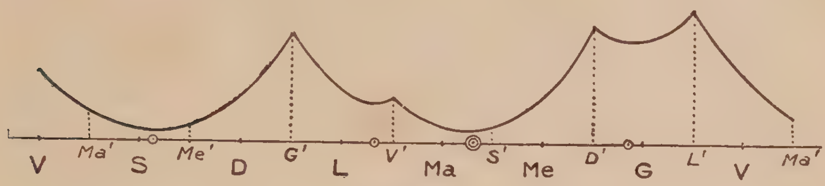

Fig. 5.

ESTENSIONE DEL CONCETTO DI MEDIA ARITMETICA IN BASE AL SUO SIGNIFICATO MECCANICO

24. — Se le modalità di una seriazione o di una serie rettilinea, che possono, come sempre, rappresentarsi mediante punti di una retta, sono $X_1\, X_2 \ldots X_s$ e ad esse corrispondono rispettivamente i valori $Y_1\, Y_2 \ldots Y_s$, allora detti $x_1\, x_2 \ldots x_s$ gli scostamenti di quelle modalità da una origine qualunque, si sa che il valore $\frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i}$, relativamente al quale nessuna ipotesi è necessario fare sulle Y_i, tranne quella che siano reali, rappresenta non soltanto la media aritmetica ponderata dei numeri X_i, ma anche l'ascissa del baricentro dei punti X_i affetti dalle masse Y_i, ossia il punto d'applicazione della risultante di più forze parallele, di intensità Y_i applicate ai punti X_i.

Il punto stesso è anche, come si sa, il centro di sospensione della retta, cioè il punto al quale si deve immaginare sospesa tale retta affinchè essa rimanga in equilibrio quando ai suoi punti X_1, $X_2 \ldots X_s$ siano applicate le masse, positive o negative $Y_1\, Y_2 \ldots Y_s$.

Passando ora ad una serie ciclica, dipendente da un carattere qualitativo X di cui diremo ancora $X_1\, X_2 \ldots X_s$ le modalità, mentre $Y_1\, Y_2 \ldots Y_s$ saranno i corrispondenti valori di una variabile Y, (*valori che supporremo ora tutti non negativi* e che potranno, in particolare, costituire le frequenze assolute o relative di quelle

modalità), viene naturale, in accordo con quella interpretazione meccanica, di rappresentare la serie ciclica:

o mediante punti della circonferenza sulla quale vengono figurate le modalità $X_1\ X_2 \ldots X_s$ di X, intendendo che ciascun punto X_i sia affetto da una massa Y_i;

oppure mediante vettori uscenti dal centro C di quella stessa circonferenza, secondo le direzioni CX_1, CX_2, $\ldots CX_s$ e coi moduli rispettivi $Y_1\ Y_2 \ldots Y_s$.

a) Conformemente alla prima di queste rappresentazioni si potrà dire, per il principio di conservazione delle leggi formali, che qualsiasi centro di sospensione della circonferenza è immagine di una modalità di X costituente una media aritmetica della data serie (1).

Se, invece, si assume l'altra rappresentazione, determinata la risultante di quel sistema di vettori, essa avrà una certa direzione che taglierà la detta circonferenza in un punto G, e un certo modulo; e il punto G sarà immagine della media aritmetica delle serie. Evidentemente una tale definizione offrirà, rispetto alla precedente, il vantaggio che (tranne nel caso in cui i vettori abbiano la somma nulla, cioè rappresentino un sistema di forze in equilibrio) il punto o modalità G sarà determinato in modo unico.

Ora, per talune serie cicliche può convenire di rappresentare le modalità di X e i rispettivi valori di Y mediante punti pesati di una circonferenza, cosicchè una loro media sarà uno dei centri di sospensione della circonferenza stessa; mentre per altre si ravviserà una maggiore opportunità ad assimilare le modalità di X e i corrispondenti valori di Y a vettori di un piano, che potranno poi sintetizzarsi mediante la loro risultante.

Si vogliano, in primo luogo, determinare i centri di sospensione di una circonferenza (giacente in un piano verticale) di cui alcuni punti $X_1\ X_2 \ldots X_s$, equispaziati, siano affetti dalle masse $Y_1\ Y_2 \ldots Y_s$. Supponiamo che i punti $X_1\ X_2 \ldots X_s$ vengano successivamente incontrati in quest'ordine quando la circonferenza sia percorsa, a partire da X_1, in senso positivo, cosicchè se, nello stesso senso $x_1 = x$ è l'ampiezza dell'arco minimo compreso fra P ed X, essendo P un

(1) Per « centro di sospensione » della circonferenza si intenderà un suo punto tale che sospendendo in esso la circonferenza, questa, per effetto dei pesi di cui è gravata, si disponga in tal modo che il suo centro (di figura) venga ad appartenere alla verticale passante per quel punto.

punto qualunque della circonferenza, le ampiezze degli archi compresi fra P e tutti i punti $X_1\ X_2 \ldots X_s$ considerati risulteranno dal seguente prospetto:

Punti della circonferenza:

$X_1 \qquad X_2 \ldots \qquad X_i \ldots \qquad X_s$

Ampiezza degli archi da P (in senso positivo):

$$x_1 = x \quad x_2 = x + \frac{2\pi}{s} \ldots x_i = x + (i-1)\frac{2\pi}{s} \ldots x_s = x + (s-1)\frac{2\pi}{s}$$

Valori di Y:

$Y_1 \qquad Y_2 \ldots \qquad Y_i \ldots \qquad Y_s$.

Affinchè P sia un centro di sospensione della circonferenza è necessario e sufficiente che la direzione della gravità sia parallela al diametro per P, e che la somma algebrica dei momenti, rispetto a tale diametro, dei pesi $Y_1\ Y_2 \ldots Y_s$ sia nulla, cioè:

$$\sum_{i=1}^{s} Y_i \, sen\, x_i = 0$$

ossia che

$$\Sigma\, Y_i \left(sen\, x \cos (i-1)\frac{2\pi}{s} + \cos x\, sen(i-1)\frac{2\pi}{s} \right) = 0$$

o più brevemente

$$A\, sen\, x + B \cos x = 0 \tag{1}$$

essendo

$$A = Y_1 \cos 0 + Y_2 \cos \frac{2\pi}{s} + \ldots + Y_i \cos (i-1)\frac{2\pi}{s} +$$
$$+ \ldots + Y_s \cos (s-1)\frac{2\pi}{s}$$

$$B = Y_1\, sen\, 0 + Y_2\, sen \frac{2\pi}{s} + \ldots + Y_i\, sen\, (i-1)\frac{2\pi}{s} +$$
$$+ \ldots + Y_s\, sen\, (s-1)\frac{2\pi}{s}.$$

La (1) fornisce senz'altro

$$tg\, x = -\frac{B}{A}$$

da cui si ricavano per l'ampiezza x valori corrispondenti a due punti della data circonferenza, diametralmente opposti fra loro, punti

che ne costituiranno i centri di sospensione e quindi le immagini di medie aritmetiche della data serie, nel senso dichiarato, se $X_1\, X_2 \ldots X_s$ sono le immagini delle modalità di X ed $Y_1\, Y_2 \ldots Y_s$ i corrispondenti valori di Y.

Va rilevato che se $A = 0$ ciò significa che è nulla la somma delle proiezioni dei vettori di direzioni OX_i e moduli Y_i sopra il diametro per X_1. Se $A = 0$, $B = 0$ è nulla la somma di tali proiezioni tanto sul diametro per X_1, quanto su quello ad esso ortogonale (e quindi è nulla l'analoga somma di proiezioni sopra un diametro qualunque). In quest'ultimo caso risulta indeterminato il valore di $tg\ x$, e qualunque punto della circonferenza è centro di sospensione. Ciò in particolare si verifica, poichè i punti X_i si sono supposti equispaziati, se $Y_1 = Y_2 = \ldots = Y_s$.

Naturalmente i centri di sospensione della data circonferenza potrebbero facilmente determinarsi anche se i punti $X_1\, X_2 \ldots X_s$ non fossero equispaziati; e quindi le medie aritmetiche, identificate con quei centri, potrebbero definirsi anche per serie cicliche non equispaziate.

b) Venendo ora all'altra forma di rappresentazione, che consiste nell'assimilare le modalità $X_1\, X_2 \ldots X_s$ e i corrispondenti valori $Y_1\, Y_2 \ldots Y_s$ a dei vettori di un piano, uscenti dal centro C della circonferenza su cui vengono figurate le $X_1\, X_2 \ldots X_s$, si tratta di determinare il punto Q della circonferenza pel quale passa la risultante dei vettori di direzioni CX_1, $CX_2 \ldots CX_s$ e di moduli $Y_1\, Y_2 \ldots Y_s$. Con le stesse notazioni di prima, salvo la sostituzione di Q a P, si avrà che, affinchè la risultante passi per Q sarà necessario che la somma delle proiezioni dei vettori sulla direzione ortogonale a CQ sia nulla, cioè che

$$\Sigma\, Y_i \cos\left(\frac{\pi}{2} - x_i\right) = \Sigma\, Y_i \operatorname{sen} x_i = 0$$

e che la somma delle proiezioni dei vettori nella direzione e nel senso di CQ sia non negativa, cioè che

$$\Sigma\, Y_i \cos x_i \geq 0.$$

Conservando ad A e B i significati già attribuiti precedentemente, queste due condizioni si potranno anche scrivere:

$$(1') \qquad A \operatorname{sen} x + B \cos x = 0$$
$$(2) \qquad A \cos x - B \operatorname{sen} x \geq 0$$

delle quali la (1′) coincide con la (1) già vista; mentre per il senso opposto a quello di CQ verrebbe:

$$\begin{aligned} &(1'') \quad A \; sen \; x' + B \; cos \; x' = 0 \\ &(2') \quad A \; cos \; x' - B \; sen \; x' \leq 0 . \end{aligned}$$

Viceversa, se fosse $A \; cos \; x - B \; sen \; x \leq 0$, nel senso opposto risulterebbe $A \; cos \; x' - B \; sen \; x' \geq 0$. Ora se $A = 0, B = 0$ le (1′) e (2) si verificheranno per qualunque valore di x, e i vettori avranno la risultante nulla, cioè saranno in equilibrio.

Se A e B non sono entrambi nulli, non potrà essere simultaneamente

$$\begin{aligned} A \; sen \; x + B \; cos \; x = 0 \\ A \; cos \; x - B \; sen \; x = 0 \end{aligned}$$

perchè di questo sistema omogeneo rispetto ad A e B, il determinante è evidentemente non nullo; e quindi determinato x (in modo unico) così che si abbia $A \; sen \; x + B \; cos \; x = 0$, si verificherà certamente una delle due condizioni $A \; cos \; x - B \; sen \; x \geq 0$, oppure $A \; cos \; x' - B \; sen \; x' \geq 0$. D'altronde la condizione (1′) è quella che già definiva i centri di sospensione della circonferenza, nella interpretazione precedente; e perciò concludiamo che il punto Q per cui passa la risultante del sistema dei vettori considerati (quando essa non sia nulla) è uno dei due centri di sospensione della circonferenza stessa.

Le due interpretazioni meccaniche considerate vengono dunque sostanzialmente a coincidere: basterà perciò, qualunque sia l'interpretazione prescelta, determinare anzitutto quel punto della circonferenza (generalmente unico) in cui la risultante dei vettori incontra la circonferenza stessa; se poi si vorranno i centri di sospensione essi saranno costituiti da tale punto e da quello diametralmente opposto.

Il procedimento analitico più semplice per determinare la risultante di più vettori complanari è quello di scomporre ciascuno di tali vettori nella somma di due, secondo direzioni arbitrarie e costanti giacenti nello stesso piano, di fare la somma algebrica di queste componenti sopra ciascuna delle due direzioni, e di trovare infine la risultante delle due somme così ottenute.

Un esempio tipico, in cui si ravvisa particolarmente l'opportunità di assimilare le modalità di X associate ai rispettivi valori di Y, a vettori di un piano e di sintetizzare il sistema mediante la risultante di tali vettori (con che la media aritmetica della serie sarà

rappresentata dal punto in cui tale risultante incontra la solita circonferenza), ci è fornito dalla serie ciclica costituita dalle direzioni X del vento che successivamente spira in un certo punto in un dato intervallo di tempo, e dalle frequenze Y relative alle singole direzioni (1).

Ciascuna delle infinite direzioni secondo cui il vento può spirare nel punto considerato è determinata dall'angolo azimutale che esso forma con una direzione fondamentale, p. es. con la Ovest-Est: ma praticamente (come si è già detto a proposito delle serie cicliche) si divide il piano orizzontale di quel punto in 8 o in 16 angoli uguali uno dei quali abbia come bisettrice la direzione fondamentale Ovest-Est, dopo di che tutte le direzioni che cadono entro ciascuno di tali angoli si assimilano alla direzione della bisettrice dell'angolo stesso, ottenendo così 8 o 16 modalità cicliche equispaziate, che costituiscono le direzioni della rosa dei venti. Se in un certo intervallo di tempo, p. es. nello spazio di un anno, si sono fatte, con ritmo regolare (di solito a 9^h, 15^h, 21^h) osservazioni anemometriche, segnando ogni volta la direzione del vento, ad ognuna delle direzioni $(X_1\ X_2 \ldots X_s)$ della rosa dei venti corrisponderà, per il complesso delle osservazioni di quell'anno, una certa frequenza $(Y_1\ Y_2 \ldots Y_s)$, e si tratterà di trovare la direzione della risultante dei vettori complanari aventi le direzioni X_i e i moduli Y_i.

Per la scomposizione dei vettori prenderemo come direzioni per il punto C in cui è posto l'osservatorio anemometrico quella *C-Est* (asse delle ascisse) e quella *C-Nord* (asse delle ordinate), di modo che le componenti su tali assi del vettore di direzione CX_i e modulo Y_i saranno:

$$Y_i \cos\ (Est \frown X_i) \quad \text{e} \quad Y_i \operatorname{sen}\ (Est \frown X_i)$$

La risultante o somma dei vettori avrà come componenti sugli stessi assi:

$$\sum_{i=1}^{s} Y_i \cos\ (Est \frown X_i) \quad \text{e} \quad \sum_{i=1}^{s} Y_i \operatorname{sen}\ (Est \frown X_i)$$

e quindi il suo modulo R sarà dato dalla:

$$R^2 = \left\{ \Sigma\, Y_i \cos\ (Est \frown X_i) \right\}^2 + \left\{ \Sigma\, Y_i \operatorname{sen}\ (Est \frown X_i) \right\}^2 ,$$

(1) In talune ricerche, specialmente in quelle che abbiano per oggetto di misurare l'effetto meccanico complessivo prodotto dal vento in un determinato punto e per un certo periodo di tempo, converrà sostituire alle semplici frequenze le corrispondenti somme delle velocità osservate.

cosicchè l'anomalia α della risultante stessa rispetto alla direzione *C-Est* sarà definita pienamente dalle

$$\cos \alpha = \frac{\Sigma\, Y_i \cos\, (Est^\frown X_i)}{R}\ ,\ sen\ \alpha = \frac{\Sigma\, Y_i\, sen\, (Est^\frown X_i)}{R}$$

La modalità o direzione definita dall'angolo α è la media aritmetica della serie ciclica considerata, secondo l'interpretazione meccanica esposta; e si può anche dire che immagine di tale modalità è il punto in cui la solita circonferenza è incontrata dalla risultante trovata.

Questo procedimento si può applicare allo studio della direzione dei venti dominanti in un punto della terra, in diversi modi, a seconda del particolare fine che si voglia conseguire.

Si può tener conto delle sole osservazioni fatte in un'ora determinata, p. es., alle 9^h, in un certo lasso di tempo, ed associare ad ognuna delle 8 o 16 direzioni convenzionali X_i la rispettiva frequenza Y_i osservata in quel periodo di tempo; ed allora la risultante segnerà la direzione del vento generalmente dominante in quell'ora. Oppure, riferendosi ancora alle stesse osservazioni, si potrà tener conto in ciascuna di esse della velocità del vento ed eseguire in corrispondenza ad ogni direzione X_i la somma Y_i delle velocità, e la direzione della risultante sarà quella in cui si è manifestato complessivamente l'effetto meccanico del vento che è spirato intorno alle ore 9 in quello stesso periodo di tempo.

Si potranno anche cumulare tutte le osservazioni fatte alle 9^h, alle 15^h, alle 21^h e tener conto o della sola frequenza o anche della velocità, ecc. Infine, disponendo di un anemografo registratore continuo, si potrà avere per ciascun giorno o la risultante delle sole direzioni o la risultante delle velocità (vettori) ed eseguire poi la somma (nel senso vettoriale) delle une o delle altre risultanti per tutto il periodo di tempo considerato (Si vedano, in fine della memoria, alcune applicazioni di tali procedimenti).

Sostituzione della risultante ad una serie ciclica

25. — Mentre la media di una seriazione o di una serie rettilinea dipendente da un carattere qualitativo X, ad s modalità, può quantitativamente sostituire l'intera seriazione o serie, qualora si intenda che a tale media corrisponda come valore di Y la somma

$\sum_{i=1}^{s} Y_i$, per le medie di una serie ciclica definita mediante la proprietà $\Sigma\varepsilon = o$ o $\Sigma\varepsilon^2 = min.$ non si potrebbe dire nulla di analogo. Se, invece, la media di una serie ciclica si definisce, come si è testè fatto, in base alla considerazione di vettori rappresentativi delle varie modalità di X e dei corrispondenti valori di Y, allora tale media si potrà, anch'essa, sostituire quantitativamente a tutta la serie, purchè le si faccia corrispondere, come valore di Y, il modulo della risultante dei vettori considerati. Perciò, tutte le volte che siano date più serie cicliche

$$\begin{matrix} Y_{11} & Y_{12} & \ldots & Y_{1s} \\ Y_{21} & Y_{22} & \ldots & Y_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ Y_{r1} & Y_{r2} & \ldots & Y_{rs} \end{matrix}$$

costituite dalle modalità di uno stesso carattere quantitativo Y e dipendenti da uno stesso carattere qualitativo X, di modalità $X_1\ X_2 \ldots X_s$ si potranno, volendo, paragonare fra di loro col sostituire a ciascuna di esse la rispettiva risultante: il paragone non dovrà vertere soltanto sulle « direzioni » di queste risultanti (direzioni che sulla solita circonferenza individuano le medie aritmetiche delle varie serie, in senso meccanico), ma anche sui moduli delle risultanti stesse.

Insistiamo nel rilevare che tutto quanto si è detto, dal n. 24 in poi, per serie cicliche equispaziate potrebbe ripetersi anche per serie cicliche non equispaziate. Nella pratica sarà tuttavia preferibile, semprechè si possa farlo, considerare modalità equispaziate, anche per il fatto che l'errore proveniente dall'assimilare tutte le modalità di un intervallo parziale ad una di esse, avrà un limite superiore costante per tutte le modalità.

Moda di una serie ciclica

26. — La consueta definizione di *moda* o *norma* è senz'altro applicabile alle serie cicliche. Se $Y = S\,((X))$ è una serie ciclica di frequenze, *moda* è una qualunque delle modalità di X alla quale corrisponde un massimo relativo della (frequenza) Y; ma, per estensione, la stessa definizione di moda si può assumere anche se Y non costituisca la frequenza di X, e sia un carattere quantitativo, a valori non negativi, comunque collegato ad X.

Mediana di una serie ciclica

27. — Anche il concetto di *mediana* si può estendere alle serie cicliche. Ricordiamo che per una seriazione o per una serie rettilinea di frequenze, mediana è una modalità X_m effettiva o di conto della variabile o della mutabile X tale che risulti (Cfr. n. 8):

$$(1) \qquad \sum_{i=1}^{m-1} Y_i < \frac{1}{2} \sum_{i=1}^{s} Y_i \,, \quad \sum_{i=1}^{m} Y_i \geq \frac{1}{2} \sum_{i=1}^{s} Y_i$$

od anche, ciò che è lo stesso:

$$(1') \qquad \sum_{i=1}^{m-1} Y_i < \sum_{i=m}^{s} Y_i \,, \quad \sum_{i=1}^{m} Y_i \geq \sum_{i=m+1}^{s} Y_i$$

il che implica l'esistenza di una prima X_1, e di un'ultima X_s modalità di X. La stessa definizione si può assumere anche se Y sia un altro carattere quantitativo, a valori positivi, comunque collegato ad X. Ma, evidentemente, tale definizione non è applicabile ad una serie ciclica, per cui non esiste una prima e non esiste un'ultima modalità; a meno che fra le modalità della serie ciclica non se ne assuma convenzionalmente una come prima e si enumerino successivamente le altre in uno dei due sensi in cui il ciclo può essere percorso. Allora, poichè una qualunque delle s modalità potrà riguardarsi come prima, la serie data ammetterà una mediana in corrispondenza ad ogni modalità considerata come prima, e quindi, in tutto, s mediane (di cui alcune potranno eventualmente coincidere).

Così se nella serie degli esempi 6º e 8º

S	*SE*	*E*	*NE*	*N*	*NW*	*W*	*SW*
3	5	0	4	2	7	3	6

si riguardano successivamente come prime modalità

S *SE* *E* *NE* *N* *NW* *W* *SW*

si trova che le corrispondenti « mediane » sono

NW *NW* *W* *W* *SW* *SW* *SE* *NE*;

e similmente, per la serie dell'esempio 9º:

L	*Ma*	*Me*	*G*	*V*	*S*	*D*
5	2	1	6	3	0	5

si vede che, a seconda che si assumano come prime modalità

$$L \quad Ma \quad Me \quad G \quad V \quad S \quad D$$

le corrispondenti mediane sono

$$G \quad V \quad D \quad D \quad L \quad Ma \quad Ma.$$

Ora è chiaro che una tale assunzione del concetto di mediana sarebbe piuttosto artificiosa, in quanto farebbe dipendere la determinazione di una mediana dal particolare modo di considerare la serie ciclica data, come iniziantesi da una piuttosto che da un'altra delle sue modalità. Per evitare tale procedimento faremo ancora ricorso al principio di conservazione delle leggi formali. Sappiamo, infatti, che nel caso delle seriazioni e delle serie rettilinee la funzione

$$\Theta(X) = \sum_{i=1}^{s} |X X_i| \, Y \tag{2}$$

e cioè la somma degli scostamenti assoluti da X delle varie modalità del ciclo, pesati tali scostamenti mediante i corrispondenti valori di Y, acquista il suo minimo valore se e soltanto se X costituisce la mediana della data seriazione o serie; e *poichè la funzione* (2) *si può evidentemente costruire anche per una serie ciclica, così si potrà dire mediana di una tale serie una modalità* $\overline{X}$ *soddisfacente alla condizione che* $\Theta(\overline{X})$ *risulti minima.*

Si tratta quindi di studiare il comportamento della $\Theta(X)$ per una serie ciclica e di individuarne i punti di minimo: ciò che ora faremo, distinguendo il caso di s pari da quello di s dispari.

Forma di $\Theta(X)$ per s pari ($s = 2k$)

28. — Adottando le stesse notazioni del n. 16, indicheremo con X_t, X_{t+1}, X_{t+2} tre modalità del ciclo, consecutive nel senso positivo; con X_{t+k}, X_{t+k+1}, X_{t+k+2} le modalità rispettivamente opposte; con Y_t, Y_{t+1}, Y_{t+2}, Y_{t+k}, Y_{t+k+1}, Y_{t+k+2} i corrispondenti valori di Y. Per la modalità opposta ad una qualunque, si potrà, come allora, adottare la stessa notazione di quest'ultima con l'apposizione di un accento; cosicchè, per la supposta parità di s sarà:

$$X_t = X'_{t+k}, \quad X_{t+1} = X'_{t+k+1}, \quad X_{t+2} = X'_{t+k+2} \text{ (1)}$$

(1) Gli indici impiegati denotano il posto delle varie modalità rispetto ad X_t, purchè si ricordi che $X_p = X_q$ se $p \equiv q$ (mod. s).

Determiniamo la forma di $\Theta(X)$ nell'intervallo $X_t \ldots X_{t+1}$ assumendo l'origine X_t. Qualunque sia la posizione di X entro $X_t \ldots X_{t+1}$, il diametro per X divide le s modalità in due sistemi di k modalità ciascuno:

$$\left.\begin{array}{l} X_{t+1}\ X_{t+2} \ldots X_{t+k} \text{ di scostamenti } 1, 2 \ldots k \\ \text{e} \\ X_t\ X_{t-1} \ldots X_{t-k+1}\ (= X_{t+k+1}) \text{ di scostamenti } 0, -1, \ldots -k+1 \end{array}\right\} \begin{array}{c}\text{dalla}\\ \text{origine}\\ X_t\end{array}$$

Perciò

$$\Theta(X) = \{|XX_{t+1}|Y_{t+1} + |XX_{t+2}|Y_{t+2} + \ldots + |XX_{t+k}|Y_{t+k}\} + \\ + \{|XX_{t+k+1}|\,Y_{t+k+1} + \ldots + |XX_{t-1}|\,Y_{t-1} + |XX_t|\,Y_t\}$$

ossia, indicando con x lo scostamento algebrico di X dall'origine X_t risulterà:

$$\Theta(X) = \{(1-x)\ Y_{t+1} + (2-x)\ Y_{t+2} + \ldots + (k-x)\ Y_{t+k}\} + \\ + \{(x+k-1)\ Y_{t+k+1} + \ldots + (x+1)\ Y_{t-1} + (x+0)\ Y_t\}$$

cioè

$$(3) \quad \Theta(X) = x\{-Y_{t+1} - Y_{t+2} \ldots - Y_{t+k} + Y_{t+k+1} + \ldots + Y_{t-1} + Y_t\} + \\ + \{1.Y_{t+1} + 2.Y_{t+2} + \ldots + k\ Y_{t+k} + (k-1)\ Y_{t+k+1} + \ldots + 1.Y_{t-1} + 0Y_t\}$$

e infine

$$(3') \qquad \Theta(X) = p_1\, x + q_1$$

dove il significato di p_1 e di q_1 è evidente. La $\Theta(X)$ è dunque una funzione lineare di X nell'intervallo $X_t \ldots X_{t+1}$ ciò che del resto è ovvio, dato che $\Theta(X)$ è, all'infuori dei fattori Y_i, una somma di scostamenti assoluti fra modalità del ciclo, e dato che ciascuno di tali scostamenti assoluti ha un decorso lineare entro lo stesso intervallo.

È ora evidente che anche in $X_{t+1} \ldots X_{t+2}$ la $\Theta(X)$ sarà una funzione lineare di x, essendo x lo scostamento di X dalla nuova origine X_{t+1}, e precisamente si avrà:

$$(4) \qquad \Theta(X) = x\{-Y_{t+2} - Y_{t+3} \ldots - Y_{t+k+1} + \\ + Y_{t+k+2} + \ldots + Y_t + Y_{t+1}\} + \\ + \{1.Y_{t+2} + 2.Y_{t+3} + \ldots + k.Y_{t+k+1} + \\ + (k-1)\ Y_{t+k+2} + \ldots + 1.Y_t + 0.Y_{t+1}\}$$

ossia

$$(4') \qquad \Theta(X) = p_2\, x + q_2 .$$

Paragonando le (3) e (4) risulta subito che:

$$(5) \qquad p_2 = p_1 + 2\ Y_{t+1} - 2\ Y_{t+k+1},$$

ossia che *nell'atto di attraversare la modalità* X_{t+1} *la funzione* $\Theta(X)$ *si mantiene lineare e continua, ma il coefficiente di* X *viene aumentato di due volte il valore di* Y *corrispondente alla modalità stessa, e diminuito di due volte il valore di* Y *corrispondente alla modalità opposta a quella attraversata.*

Il paragone delle (3) e (4) fornirebbe anche la relazione fra q_1 e q_2: ma il modo più semplice di determinazione di q_2 si ha dall'osservare che, essendo la $\Theta(X)$ continua in X_{t+1} ed avendo assunto in $X_t \ldots X_{t+1}$ l'origine X_t, ed in $X_{t+1} \ldots X_{t+2}$ l'origine X_{t+1}, si dovrà avere:

$$(p_1 x + q_1)_{x=1} = (p_2 x + q_2)_{x=0}$$

ossia

$$q_2 = p_1 + q_1, \tag{6}$$

vale a dire *attraversando* X_{t+1} *il termine noto di* $\Theta(X)$ *si accresce di una quantità uguale al coefficiente di* x (1).

Calcolati, perciò, mediante la (3), i coefficienti p_1 *e* q_1 della $\Theta(X)$ in $X_t \ldots X_{t+1}$ rispetto all'origine X_t i coefficienti analoghi che determineranno via via la $\Theta(X)$ negli intervalli successivi a quel primo, coi rispettivi riferimenti alle origini $X_{t+1}, X_{t+2}, \ldots, X_{t-1}$, si avranno ricorrentemente mediante le (5) e (6) ottenendo, così, la espressione di $\Theta(X)$ in tutto il ciclo. Come riprova della esattezza dei calcoli, si potranno ulteriormente applicare le (5) e (6), e si dovrà ritrovare la $\Theta(X)$ in $X_t \ldots X_{t+1}$.

La rappresentazione geometrica di $\Theta(X)$ in un sistema cartesiano nel quale vengano, sull'asse delle ascisse, rappresentate le modalità di X, come già si fece ai nn. 19 e 23, è quindi costituita da una poligonale i cui s lati hanno per proiezioni sull'asse delle ascisse i successivi segmenti come $X_t \ldots X_{t+1}$ (*Vedi* grafico in fine di questo numero).

L'ordinata minima di tale spezzata, che dovrà necessariamente corrispondere ad un vertice della spezzata stessa (ossia il minimo valore di $\Theta(X)$), *si avrà quindi per una modalità di X che si potrà considerare come mediana della serie ciclica.* In questo senso, non è perciò esclusa la possibilità dell'esistenza di più mediane per una medesima serie ciclica: in particolare, se in due modalità consecutive

(1) A differenza di quanto si è fatto ai nn. 17 e 22 per la rappresentazione di $\mu_1(X)$ e di $\mu_2(X)$ per cui conveniva conservare, lungo tutto il ciclo, una stessa origine, è preferibile per la $\theta(X)$ cambiare origine, come qui si è fatto, da intervallo a intervallo.

del ciclo, p. es. in X_t ed X_{t+1} la $\Theta(X)$ raggiunge il suo minimo valore, il lato della spezzata sovrastante ad $X_t \ldots X_{t+1}$ sarà parallelo all'asse delle ascisse, e qualunque modalità (di conto) compresa fra X_t ed X_{t+1} si potrà considerare come una mediana della serie data (1).

Si osservi che nell'intervallo $X'_t \ldots X'_{t+1}$ opposto ad $X_t \ldots X_{t+1}$ la $\Theta(X)$ assume la forma (origine X'_t):

$$\Theta(X) = Px + Q$$

dove, per la (5) e per la (3), è:

$$P = p_1 + 2(Y_{t+1} + Y_{t+2} + \ldots + Y_{t+k}) -$$
$$- 2(Y_{t+k+1} + \ldots + Y_{t-1} + Y_t) = - p_1,$$

mentre in $X_t \ldots X_{t+1}$ era $\Theta(X) = p_1 x + q_1$, cosicchè nella poligonale rappresentativa di $\Theta(X)$ due lati corrispondenti ad intervalli opposti hanno coefficienti angolari tra loro contrarî. Segue di qui che se la $\Theta(X)$ ha in X_i un massimo, in X'_i avrà un minimo.

Esempio 11º. — Applichiamo l'esposto metodo di determinazione della $\Theta(X)$ alla serie ciclica degli esempi 6º ed 8º (n. 19) così costituita:

Modalità di X:	S	SE	E	NE	N	NW	W	SW
Valori di Y:	3	5	0	4	2	7	3	6

(1) È ovvia l'analogia con il caso di indeterminazione della mediana di un sistema di $2k$ numeri.

INTERVALLI	ORIGINE	$\Theta\,(X)$ (1)	MODALITÀ attraversata e relativo valore di Y (2)	MODALITÀ opposta e relativo valore di Y (3)	FORMA DI $\Theta\,(X)$ (4)	$q_i = (1)_{x=1}$ (5)
S ... SE	S	$8x + 58$	..	..	..	..
		..	SE, 5	$N'W'$, 7	$(8 + 10 - 14)\,x + q_2$	$q_2 = 8 + 58$
SE ... E	SE	$4x + 66$	..	..	..	..
		..	E, 0	W', 3	$(4 + 0 - 6)\,x + q_3$	$q_3 = 4 + 66$
E ... NE	E	$-2x + 70$	..	..	..	..
		..	NE, 4	$S'W'$, 6	$(-2 + 8 - 12)\,x + q_4$	$q_4 = -2 + 70$
NE ... N	NE	$-6x + 68$	..	..	..	..
		..	N, 2	S', 3	$(-6 + 4 - 6)\,x + q_5$	$q_5 = -6 + 68$
N ... NW	N	$-8x + 62$	..	..	..	..
		..	NW, 7	$S'E'$, 5	$(-8 + 14 - 10)\,x + q_6$	$q_6 = -8 + 62$
NW ... W	NW	$-4x + 54$	..	..	..	..
		..	W, 3	E', 0	$(-4 + 6 - 0)\,x + q_7$	$q_7 = -4 + 54$
W ... SW	W	$2x + 50$	..	..	..	..
		..	SW, 6	$N'E'$, 4	$(2 + 12 - 8)\,x + q_8$	$q_8 = 2 + 50$
SW ... S	SW	$6x + 52$	..	..	..	..
		..	S, 3	N', 2	$(6 + 6 - 4)\,x + q_1$	$q_1 = 6 + 52$
PROVA: S ... SE	S	$8x + 58$	..	..	..	..

La colonna (1) contiene le espressioni di $\Theta(X)$ nei successivi intervalli. Nel primo di questi, $S \ldots SE$, la $\Theta(X) = 8x + 58$ è calcolata direttamente mediante la (3). Nella colonna (2) sono indicate le modalità che si attraversano nel passaggio da ciascun intervallo al consecutivo, modalità che coincidono con le opposte di altre modalità, come risulta dalla colonna (3). La (4) mostra la formazione dei successivi coefficienti di x, e la (5) dà i valori del termine q_i indipendente da x. La $\Theta(X)$ assume il valore minimo, uguale a 50, in W; e perciò W è da riguardarsi come mediana della serie ciclica proposta. L'andamento di $\Theta(X)$ è graficamente rappresentato qui presso.

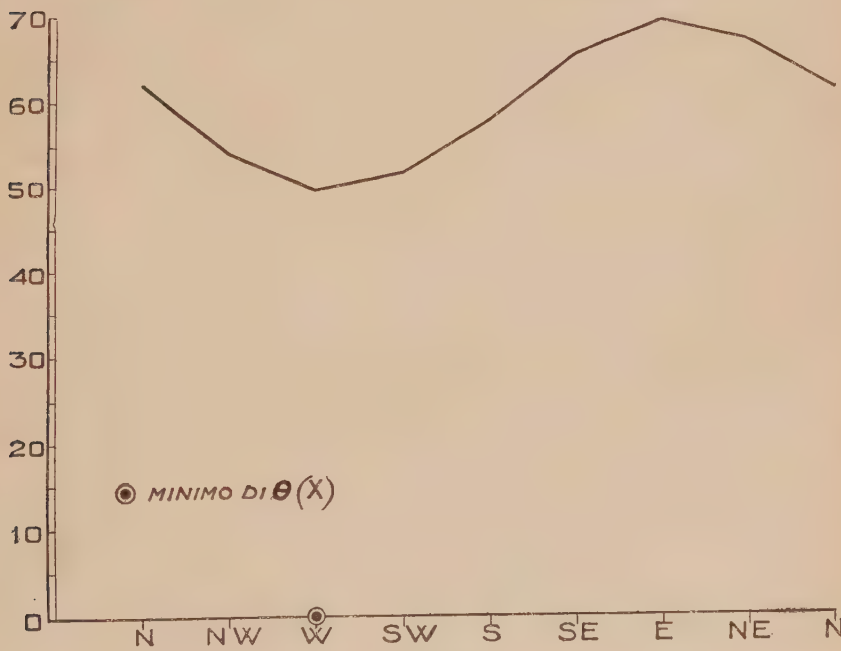

Fig. 6.

FORMA DI $\Theta(X)$ PER s DISPARI $(s = 2k + 1)$

29. — Dette ancora X_t, X_{t+1}, X_{t+2}, tre modalità consecutive nel senso positivo, per la supposta disparità di s, fra X_t ed X_{t+1} cadrà la modalità di conto X'_{t+k+1} opposta alla X_{t+k+1} e fra X_{t+1} e X_{t+2} cadrà la modalità di conto X'_{t+k+2}, opposta alla X_{t+k+2}.

Si intende immediatamente che la $\Theta(X)$ avrà in ciascuno degli intervalli $X_t \ldots X'_{t+k+1}$, $X'_{t+k+1} \ldots X_{t+1}$, $X_{t+1} \ldots X'_{t+k+2}$, $X'_{t+k+2} \ldots X_{t+2}$ e negli analoghi successivi, un decorso lineare, come somma di funzioni lineari negli stessi intervalli. Basterà dunque calcolare direttamente $\Theta(X)$ per X giacente nel primo $X_t \ldots X'_{t+k+1}$ di tali intervalli, e vedere come si modifica la forma di $\Theta(X)$: *a*) nell'atto di attraversare la modalità di conto X'_{t+k+1}; e successivamente: *b*) nell'atto di attraversare la modalità X_{t+1}.

Prenderemo come origine, per la misura degli scostamenti di una X generica: fino ad X'_{t+k+1} la modalità X_t; da X'_{t+k+1} fino ad X'_{t+k+2} l'origine X_{t+1}, e così via.

a) Per X in $X_t \ldots X'_{t+k+1}$, indicando con x lo scostamento di X e notando che gli scostamenti delle altre modalità da X_t sono:

$$\left.\begin{array}{l} \text{per} \quad X_{t+1}\ X_{t+2} \ldots X_{t+k} \\ \text{e per} \ X_{t-1}\ X_{t-2} \ldots X_{t-k} \end{array}\right\} \text{ rispettivamente } \left\{\begin{array}{rrr} 1, & 2, \ldots & k \\ -1, & -2, \ldots & -k \end{array}\right.$$

si ha:

$$\Theta(X) = \{ |XX_{t+1}|\, Y_{t+1} + \ldots + |XX_{t+k}|\, Y_{t+k} \} + \\ + \{ |XX_{t-k}|\, Y_{t-k} + \ldots + |XX_{t-1}|\, Y_{t-1} + |XX_t|\, Y_t \}$$

cioè

$$\Theta(X) = x \{ -Y_{t+1} - Y_{t+2} \ldots - Y_{t+k} + Y_{t-k} + \ldots + Y_t \} + \\ + \{ 1.Y_{t+1} + 2.Y_{t+2} \ldots + k\, Y_{t+1} + k\, Y_{t-k} + \ldots + 1.Y_{t-1} + 0.Y_t \}$$

e infine, brevemente,

$$\Theta(X) = U_1\, x + V_1$$

Se ora si passa all'intervallo $X'_{t+k+1} \ldots X_{t+1}$ e si ripete l'analogo calcolo di $\Theta(X)$, assumendo come origine X_{t+1}, si trova

$$\Theta(X) = U_2\, x + V_2\,,$$

con $U_2 = -Y_{t+1} - Y_{t+2} \ldots - Y_{t+k} - Y_{t-k} + Y_{t-k+1} + \ldots + Y_t$

e perciò

(1) $U_2 = U_1 - 2\, Y_{t-k} = U_1 - 2\, Y_{t+k+1}$ (essendo $t - k \equiv t + k + 1$, *mod.* $2k + 1$).

Quanto a V_2, si potrà rapidamente calcolarlo, in quanto che la $\Theta(X)$ è una funzione continua ovunque, e in particolare in X'_{t+k+1}, e

quindi, tenendo conto del mutamento di origine nell'attraversare X'_{t+k+1}, si dovrà avere:

$$(U_1 x + V_1)_{x=1/2} = (U_2 x + V_2)_{x=-1/2}$$

ossia

(2) $$V_2 = V_1 + U_1 - Y_{t+k+1}$$

Le (1) e (2) significano che, *nell'atto di attraversare la modalità di conto* X'_{t+k+1} *opposta alla* X_{t+k+1} *la funzione* $\Theta(X)$ *si mantiene lineare e continua ed il coefficiente di* x *diminuisce di due volte il valore di* Y *corrispondente ad* X_{t+k+1}, *mentre il termine noto diminuisce di una volta il valore stesso e cresce di una quantità uguale al coefficiente di* x.

b) Mantenendo ora l'origine X_{t+1}, e calcolando l'espressione di $\Theta(X)$ in $X_{t+1} \ldots X'_{t+k+2}$ si troverà similmente

$$\Theta(X) = U_3 x + V_3$$

con $U_3 = + Y_{t+1} - Y_{t+2} - \ldots - Y_{t+k} - Y_{t-k} + Y_{t-k+1} + \ldots + Y_t$

ossia

(3) $$U_3 = U_2 + 2\,Y_{t+1};$$

inoltre si avrà subito che

$$(U_2 x + V_2)_{x=0} = (U_3 x + V_3)_{x=0}$$

donde

(4) $$V_3 = V_2;$$

perciò, per le (3) e (4), *nell'attraversare la modalità* X_{t+1}, *nella funzione* $\Theta(X)$ *il coefficiente di* x *aumenta di due volte il valore di* Y *corrispondente a quella modalità, ed il termine indipendente da* x *rimane immutato.*

Anche in questo caso, perciò, a partire dall'espressione di $\Theta(X)$ in $X_t \ldots X'_{t+k+1}$ si potranno avere con semplici operazioni ricorrenti ed alternative le espressioni di $\Theta(X)$ in tutti i successivi intervalli del ciclo; la rappresentazione cartesiana di $\Theta(X)$ sarà ancora costituita da una poligonale, non più di s, ma di $2s$ lati (*Vedi* grafico in fine di questo numero).

Il valore minimo di $\Theta(X)$ dovrà aversi in corrispondenza ad un vertice di tale poligonale, senonchè si può escludere a priori che ciò possa accadere per quegli s vertici della poligonale che corrispondono alle modalità di conto opposte alle modalità effettive. Difatti, con-

siderati i due intervalli $X_t \ldots X'_{t+k+1}$ ed $X'_{t+k+1} \ldots X_{t+1}$, sappiamo che nel primo di essi si ha

$$\Theta(X) = U_1 x + V_1 \quad (\text{origine } X_t)$$

e quindi

$$\Theta(X_t) = \Theta(0) = V_1$$
$$\Theta(X'_{t+k+1}) = \Theta(^1/_2) = {}^1/_2 U_1 + V_1 ;$$

mentre in $X'_{t+k+1} \ldots X_{t+1}$ si ha

$$\Theta(X) = U_2 x + V_2 = (U_1 - 2 Y_{t+k+1}) x + \\ + (V_1 + U_1 - Y_{t+k+1}) \quad (\text{orig. } X_{t+1})$$

da cui

$$\Theta(X_{t+1}) = \Theta(0) = V_1 + U_1 - Y_{t+k+1} ;$$

si vede subito che se $U_1 > 0$

$\Theta(X'_{t+k+1}) > \Theta(X_t)$ (senza escludere che possa anche essere $\Theta(X'_{t+k+1}) > \Theta(X_{t+1})$;

mentre se $U_1 < 0$ è certamente

$$\Theta(X'_{t+k+1}) > \Theta(X_{t+1}) ,$$

e se $U_1 = 0$ è $\Theta(X'_{t+k+1}) = \Theta(X_t)$.

Dunque il valore di Θ nel punto X'_{t+k+1} opposto ad una modalità è sempre non minore del valore di Θ in una almeno delle due modalità limitrofe X_t ed X_{t+1}; cioè in X'_{t+k+1} la $\Theta(X)$ o non ha un minimo, oppure ha un minimo uguale a quello che eventualmente si verifica per X_t.

Si conclude che, anche nel caso di s dispari, le sole modalità fra cui basta cercare i minimi di $\Theta(X)$ sono le modalità effettive.

Infine, anche qui si verifica che due lati della spezzata corrispondenti a due intervalli opposti del ciclo hanno coefficienti angolari fra loro contrarî; onde segue che se in una modalità X_i la $\Theta(X)$ ha un minimo, nella modalità opposta X'_i essa avrà un massimo.

Esempio 12°. — Applichiamo le cose dette alla serie ciclica dell'esempio 9° così costituita:

Modalità di X:	L	Ma	Me	G	V	S	D	$(s = 7)$
Valori di Y:	5	2	1	6	3	0	5	

Si calcola direttamente la forma di $\Theta(X)$ nell'intervallo $L\,V'$, e si trova $\Theta(X) = 4x + 36$. Le espressioni di $\Theta(X)$ nei 13 intervalli successivi si determinano con le semplici operazioni ricorrenti sopra descritte, come risulta dal prospetto seguente:

Intervallo	Origine	$\Theta(X) =$ (1)	Modalità attrav. e relativo valore di Y (2)	Modalità opposta e relativo valore di Y (3)	Forma di $\Theta(X)$ (4)	$V_j = V_{j-1} + U_{j-1} - (3)$ (5)
$L\ V'$	L	$4x + 36$	..	..	..	..
		..	..	V', 3	$(4-6)x + V_2$	$V_2 = 4 + 36 - 3 = 37$
$V'\ Ma$	Ma	$-2x + 37$	..	..	..	..
		..	Ma, 2	..	$(-2+4)x + 37$	..
$Ma\ S'$	Ma	$2x + 37$	..	..	..	..
		..	..	S', 0	$(2-0)x + V_4$	$V_4 = 2 + 37 - 0 = 39$
$S'\ Me$	Me	$2x + 39$	..	..	..	..
		..	Me, 1	..	$(2+2)x + 39$	..
$Me\ D'$	Me	$4x + 39$	..	..	..	..
		..	..	D', 5	$(4-10)x + V_6$	$V_6 = 4 + 39 - 5 = 38$
$D'\ G$	G	$-6x + 38$	..	..	..	..
		..	G, 6	..	$(-6+12)x + 38$	..
$G\ L'$	G	$6x + 38$	..	..	..	..
		..	..	L', 5	$(6-10)x + V_8$	$V_8 = 6 + 38 - 5 = 39$
$L'\ V$	V	$-4x + 39$	..	..	..	..
		..	V, 3	..	$(-4+6)x + 39$	..
$V\ Ma'$	V	$2x + 39$	..	..	..	..
		..	..	Ma', 2	$(2-4)x + V_{10}$	$V_{10} = 2 + 39 - 2 = 39$
$Ma'\ S$	S	$-2x + 39$	..	..	..	..
		..	S, 0	..	$(-2+0)x + 39$	..
$S\ Me'$	S	$-2x + 39$	..	..	..	..
		..	..	Me', 1	$(-2-2)x + V_{12}$	$V_{12} = -2 + 39 - 1 = 36$
$Me'\ D$	D	$-4x + 36$	..	..	..	..
		..	D, 5	..	$(-4+10)x + 36$	..
$D\ G'$	D	$6x + 36$	..	..	..	..
		..	..	G', 6	$(6-12)x + V_{14}$	$V_{14} = 6 + 36 - 6 = 36$
$G'\ L$	L	$-6x + 36$	..	..	..	..
		..	L, 5	..	$(-6+10)x + 36$	..
Prova: $L\ V'$	L	$4x + 36$	..	..	..	..

La $\Theta(X)$ raggiunge il suo valore minimo, 36, sia per la modalità D che per la modalità L, ciascuna delle quali potrà, pertanto, riguardarsi come una mediana del dato ciclo.

Viene qui di seguito data la rappresentazione grafica di $\Theta(X)$:

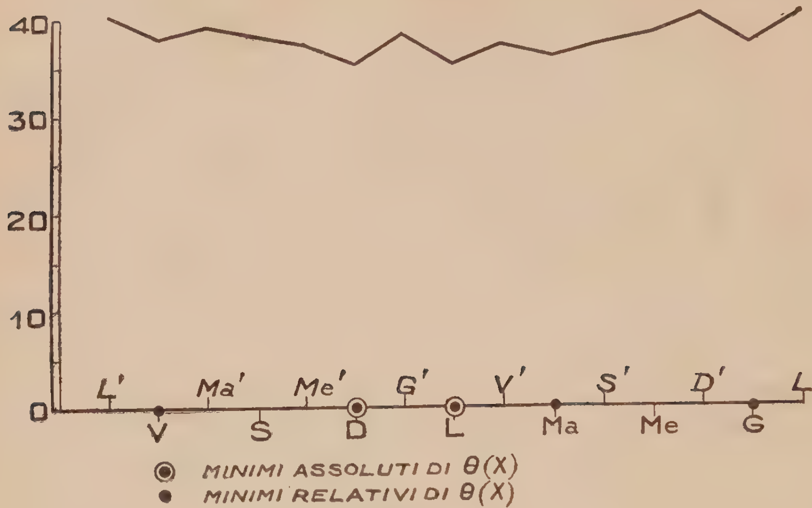

Fig. 7.

Proprietà di $\Theta(X)$

30. — Abbiamo veduto che, a seconda che sia s pari oppure dispari, la poligonale rappresentativa di $\Theta(X)$, cioè della somma degli scostamenti assoluti, è costituita da s o da $2s$ lati, e che i minimi di $\Theta(x)$ possono verificarsi soltanto in corrispondenza a quei vertici della poligonale che si proiettano in modalità effettive del ciclo, a meno che in due di essi consecutivi la $\Theta(X)$ non raggiunga uno stesso valore minimo, nel qual caso, se s è pari, tutte le modalità di conto comprese fra quelle corrispondenti a quei due vertici possono indifferentemente considerarsi come mediane della data serie.

In particolare, nell'ipotesi di s pari, se a tutte le modalità di X corrispondono uguali valori di Y, $\Theta(X)$ sarà costante e qualunque modalità di X, effettiva o fittizia, potrà riguardarsi come mediana del ciclo.

Se s è dispari e a tutte le modalità di X corrispondono uguali valori di Y, $\Theta(X)$ raggiungerà uno stesso valore minimo in corri-

spondenza ad ogni modalità effettiva, mentre nelle modalità a queste opposte raggiungerà un valore massimo costante; di modo che le mediane della data serie saranno costituite da tutte e dalle sole modalità effettive di X.

Se rispetto ad un punto della circonferenza (e quindi anche rispetto all'opposto) i punti modalità e i corrispondenti valori di Y sono disposti simmetricamente, in quel punto, oppure nel suo opposto, $\Theta(X)$ raggiunge un minimo.

Due serie cicliche aventi le stesse modalità e i valori di Y proporzionali nelle modalità corrispondenti hanno le stesse mediane.

Osserviamo infine che, mentre nel caso delle seriazioni e delle serie rettilinee la $\Theta(X)$ ammette un solo valore minimo e quindi esiste una sola mediana o, per lo meno, le infinite mediane sono costituite da tutte le modalità di un solo intervallo, nel caso delle serie cicliche la $\Theta(X)$ può ammettere più valori minimi, assoluti e relativi. È perciò necessario fare una distinzione relativamente alle modalità in cui si verificano questi minimi e cioè dire « mediane della serie data, in senso stretto » o, semplicemente, « mediane » le sole modalità di X per le quali $\Theta(X)$ acquista il valore minimo assoluto, e « mediane in senso lato » tutte le modalità di X per cui $\Theta(X)$ acquisti un valore minimo, assoluto o relativo.

Le mediane di una serie ciclica, considerate in questo senso più generale, ottemperano tutte ad una condizione *analoga* alla (1') del numero 27, nel senso che, se X_m è una di tali mediane, esistono precedentemente ad X_m, nel senso positivo, alcune modalità $X_{m-1}, X_{m-2}, \ldots X_{m-t}$ in numero conveniente, tali che

$$\sum_{i=m-t}^{m} Y_i < Y_m + Y_{m+1} + \ldots Y_{m-t-1}$$

mentre

$$\sum_{i=m-t}^{m} Y_i \geq Y_{m+1} + \ldots + Y_{m-t-1}\,.$$

Tuttavia, nei riguardi della mutabilità della serie converrà, come sarà fatto più oltre, considerare soltanto le mediane in senso stretto.

Osservazione. — Dal momento che le mediane di una serie ciclica possono esclusivamente cadere nei punti modalità, per la loro determinazione pratica ci si può limitare al calcolo di $\Theta(X)$ in questi soli punti e alla successiva scelta dei punti di minimo. È tuttavia soltanto in seguito all'osservazione del decorso di $\Theta(X)$ in tutto il

ciclo che si è potuta localizzare la posizione delle mediane; e secondariamente è importante notare che la conoscenza di $\Theta(X)$ lungo l'intero ciclo permette di determinare rapidamente lo scostamento semplice medio della serie rispetto a ciascuna delle sue medie aritmetiche (ed anche, eventualmente, rispetto ad altre modalità): difatti lo scostamento semplice medio della data serie rispetto ad una modalità $\overline{X}$ è evidentemente:

$$ {}^{1}\eta_{\overline{X}} = \frac{\Theta(\overline{X})}{\Sigma Y_i} .$$

Proprietà delle medie

31. — Premettiamo una semplice osservazione. Abbiamo visto (nn. 17 e 18) che se la circonferenza su cui si rappresentano le modalità di X si divide mediante i punti $X_1' \ldots X_t' \; X'_{t+1} \ldots X_p'$ opposti ai punti-modalità pei quali non sia nullo il rispettivo valore di Y, in corrispondenza a ciascuno degli intervalli così ottenuti, come $X_t' \ldots X'_{t+1}$ esiste una media aritmetica che è ordinaria, limite o fittizia secondo che cada internamente o in un estremo o esternamente all'intervallo. Oltre a queste possono poi esistere anche delle medie *sui generis*. Ora, se invece di considerare quei soli punti di divisione, la circonferenza venisse spartita in s intervalli mediante i punti opposti a tutti i punti modalità (effettivi o concomitanti) di X, e, poniamo col primo metodo, si applicasse il procedimento per la ricerca della media in corrispondenza a ciascuno di tali intervalli, assumendo ogni volta una origine conveniente, p. es., il punto centrale dell'intervallo, allora per ogni origine, cioè per ogni intervallo, il calcolo verrebbe a fornire lo scostamento di una certa media delle quali, tuttavia, *le sole distinte sarebbero quelle corrispondenti ai* p *intervalli prima considerati*. Così se in una data serie ciclica, alle modalità

$$X_1 \, X_2 \, X_3 \, X_4 \ldots X_s$$

corrispondessero, come valori di Y:

$$Y_1 \neq 0 , \; Y_2 = 0 , \; Y_3 = 0, \; Y_4 \neq 0 , \ldots \, Y_s ,$$

a tutto l'intervallo $X_1' \ldots X_4'$ competerebbe, come sappiamo, una sola media aritmetica A ordinaria, limite o fittizia. Se, invece di $X_1' \ldots X_4'$ si considerassero i tre intervalli $X_1' X_2'$, $X_2' X_3'$, $X_3' X_4'$ e si applicasse il primo metodo di ricerca delle medie a ciascuno di

essi, prendendo come origini i loro punti di mezzo, si avrebbero come risultati tre scostamenti (differenti di un grado ciascuno dal successivo e) designanti sempre la stessa media A che competeva a tutto l'intervallo.

Circa le denominazioni da usare per indicare le medie, in riferimento a tutti gli s detti intervalli del ciclo, è chiaro che esse saranno quelle stesse stabilite in riferimento ai soli p intervalli considerati precedentemente. Così se, per fissare le idee, la media A fosse esterna ad $X_1' \ldots X_4'$ e quindi fittizia, essa si dovrebbe pure dire fittizia rispetto ad $X_1' X_2'$, ad $X_2' X_3'$, ad $X_3' X_4'$; se, invece, cadesse internamente ad $X_1' X_4'$, p. es., in $X_3' X_4'$ e quindi costituisse una media ordinaria, allora, come competente ad $X_1' X_2'$ ed anche a $X_2' X_3'$ si dovrebbe dire fittizia, e come competente ad $X_3' X_4'$ sarebbe una media ordinaria.

In questa nuova forma di riferimento delle medie, si può dunque dire che a ciascuno degli s intervalli determinati dai punti opposti a tutti i punti modalità compete una media: ma non si può escludere che a due o più di questi intervalli competa una stessa media, e ciò si verifica quando questi certi intervalli costituiscano complessivamente uno solo di quei p intervalli che si erano finora considerati.

A ulteriore chiarimento di questo concetto, si osservi il grafico annesso all'esempio 8° del numero 19. Il ciclo (di 8 modalità) è stato diviso in 7 intervalli (perchè alla modalità W corrisponde il valore 0) a ciascuno dei quali compete una sola media, delle quali cinque ordinarie, una limite ed una fittizia. Se il ciclo si considera invece diviso negli 8 intervalli determinati dai punti opposti a tutte le modalità, l'intervallo $S'E' \ldots N'E'$ si scinde nei due $S'E' \ldots E'$ ed $E' \ldots N'E'$, e l'unica media T (ordinaria) corrispondente a $S'E' \ldots N'E'$ deve considerarsi competente sia all'intervallo $S'E' \ldots E'$ (come media fittizia), sia all'intervallo $E' \ldots N'E'$ come media ordinaria.

Nelle considerazioni che stiamo per fare si riguarderà sempre il ciclo diviso in tanti intervalli quante sono le modalità effettive e concomitanti di X, e si immaginerà associata a ciascuno di tali intervalli la media che gli compete (intendendo sempre che i punti di divisione siano quelli opposti ai punti modalità).

32. — Data una serie ciclica (equispaziata) (di s modalità $X_1\, X_2\,, \ldots X_s$) *la somma algebrica dei momenti primi calcolati nei punti modalità* (se s è dispari) *oppure nei punti di mezzo fra i successivi punti-modalità* (se s è pari), *è nulla.*

Difatti se s è dispari si ha:

$$\mu_1(X_1) = (Y_2 - Y_s) + 2(Y_3 - Y_{s-1}) + 3(Y_4 - Y_{s-2}) +$$
$$+ \ldots + \frac{s-1}{2}\left(Y_{\frac{s+1}{2}} - Y_{\frac{s+3}{2}}\right).$$

Eseguendo le successive permutazioni circolari sugli indici 1, 2 ... $\frac{s+1}{2}$, $\frac{s+3}{2}$, ... s, e addizionando membro a membro risulterà senz'altro $\sum_{i=1}^{s} \mu_1(X_i) = 0$.

Se s è pari ed $O_1\, O_2 \ldots O_s$ sono i punti di mezzo fra i punti modalità X_1 e X_2, X_2 e X_3, ..., X_s e X_1, si ha similmente:

$$\mu_1(O_1) = \frac{1}{2}(Y_1 - Y_s) + \frac{3}{2}(Y_2 - Y_{s-1}) +$$
$$+ \ldots + \frac{s-1}{2}\left(Y_{\frac{s}{2}} - Y_{\frac{s+2}{2}}\right)$$

ed anche qui, eseguendo le solite permutazioni circolari e addizionando si avrà $\sum_{i=1}^{s} \mu_1(O_i) = 0$.

Da ciò si traggono due semplici conseguenze:

a) Poichè, sia nel caso di s dispari che pari, la determinazione della media competente a ciascuno dei successivi intervalli determinati dai punti opposti ai punti modalità si può, col primo metodo di calcolo (Cfr. n. 19), eseguire assumendo volta a volta come origine O_i il punto modalità situato a metà dell'intervallo considerato (origine che coincide o no con un punto modalità secondo che sia s dispari o pari) e poichè, d'altra parte, lo scostamento di una tale media dalla rispettiva origine è (Cfr. n. 17) $\frac{\mu_1(O_i)}{\mu_0}$, così si conclude che

$$\sum_{i=1}^{s} \frac{\mu_1(O_i)}{\mu_0} = 0$$

intendendo la sommatoria estesa alle origini assunte nei successivi intervalli. Si può dunque dire: *la somma degli scostamenti delle medie ordinarie, limite e fittizie dai punti di mezzo degli intervalli ai quali competono è nulla.*

Per esempio, si prenda (dal n. 57, più oltre,) la serie dei numeri di matrimoni celebrati in Roma nei vari giorni della settimana

del periodo 1877-86, proporzionalmente a 10.000; gli scostamenti delle medie aritmetiche, ordinarie e fittizie, dai punti di mezzo dei rispettivi intervalli sono:

$x_{I} = 0{,}1295$; $\bar{x}_{II} = x_{II} - 1 = -0{,}8495$; $\bar{x}_{III} = x_{III} - 2 = -0{,}0652$; $\bar{x}_{IV} = x_{IV} - 3 = 1{,}1951$; $\bar{x}_{V} = x_{V} - 4 = 0{,}6214$; $\bar{x}_{VI} = x_{VI} - 5 = -0{,}3457$; $\bar{x}_{VII} = x_{VII} - 6 = -0{,}6856$ e risulta appunto:

$$x_{I} + \bar{x}_{II} + \bar{x}_{III} + \bar{x}_{IV} + \bar{x}_{V} + \bar{x}_{VI} + \bar{x}_{VII} = 0.$$

Così pure, si riprenda la serie dell'esempio 8 al numero 19, in cui il ciclo, a 8 modalità, è stato diviso in 7 intervalli; dividendolo, invece, in 8 intervalli mediante i punti opposti agli 8 punti modalità, agli intervalli *NW...W* ed *W...SW* corrisponde quell'unica media che si è designata con x_{VI} (e che corrispondeva a *NW...SW*); considerata quindi tale x_{VI} come competente ad *W...SW* il suo scostamento dal punto di mezzo di questo intervallo sarà $\bar{x}_{VI} = x_{VI} - 5$; considerata come competente ad *W...SW* lo scostamento dal relativo punto di mezzo sarà $\bar{x}_{VI}$-*bis* $= x_{VI} - 6$. Perciò gli scostamenti delle medie dai punti di mezzo di tutti gli 8 intervalli sono:

$$x_{I} = -\frac{23}{30} \quad \text{(fitt.)} \quad (F)$$

$$\bar{x}_{II} = x_{II} - 1 = \frac{33}{30} - 1 = \frac{3}{30} \quad \text{(ord.)} \quad (P)$$

$$\bar{x}_{III} = x_{III} - 2 = \frac{57}{30} - 2 = -\frac{3}{30} \quad \text{(ord.)} \quad (Q)$$

$$\bar{x}_{IV} = x_{IV} - 3 = \frac{105}{30} - 3 = \frac{15}{30} \quad \text{(lim.)} \quad (N = S')$$

$$\bar{x}_{V} = x_{V} - 4 = \frac{129}{30} - 4 = \frac{9}{30} \quad \text{(ord.)} \quad (R)$$

$$\bar{x}_{VI} = x_{VI} - 5 = \frac{169}{30} - 5 = \frac{19}{30} \quad \text{(fitt. riferita a } NW\ldots W\text{)} \quad (T)$$

$$\bar{x}_{IV}\text{-}bis = x_{IV} - 6 = -\frac{11}{30} \quad \text{(ord. riferita ad } W\ldots SW\text{)} \quad (T)$$

$$\bar{x}_{VII} = x_{VII} - 7 = \frac{201}{30} - 7 = -\frac{9}{30} \quad \text{(ord.)} \quad (U)$$

e risulta appunto:

$$-\frac{23}{30} + \frac{3}{30} - \frac{3}{30} + \frac{15}{30} + \frac{9}{30} + \frac{19}{30} - \frac{11}{30} - \frac{9}{30} = 0$$

b) La proprietà trovata si presta a una facile interpretazione geometrica. Poichè i momenti primi calcolati nei punti di mezzo degli intervalli determinati dai punti opposti ai punti modalità, sono evidentemente le semisomme delle basi dei trapezi di uguale altezza in cui l'area compresa fra l'asse delle ascisse e la linea rappresentativa del momento primo rimane scomposta dalle ordinate condotte pei punti di separazione di quegli intervalli, così la somma algebrica delle aree di quei trapezi è nulla, cioè la linea rappresentativa del momento primo racchiude, insieme con l'asse delle ascisse, un'area algebricamente nulla.

33. — Più in generale: se i momenti primi calcolati nei punti modalità (quando sia s dispari) oppure nei punti di mezzo fra i successivi punti modalità (quando sia s pari) si moltiplicano, ciascuno per la somma dei valori di Y corrispondenti alle due modalità aventi in valore assoluto un certo scostamento costante da quei punti, la somma dei prodotti così ottenuti è nulla, ossia, con le stesse notazioni di prima:

$$\text{per } s \text{ dispari: } \sum_{i=1}^{s} \mu_1 (X_i) \left[Y_{i+t} + Y_{i-t} \right] = 0$$

$$\text{e per } s \text{ pari: } \sum_{i=1}^{s} \mu_1 (O_i) \left[Y_{i+t+1} + Y_{i-t} \right] = 0.$$

Anche qui la dimostrazione ha il carattere di una semplice prova come nel caso precedente.

In particolare si deduce che, nel caso di s dispari, se i momenti primi calcolati nei punti modalità si moltiplicano pei valori di Y corrispondenti alle modalità stesse, la somma di tali prodotti è nulla, cioè:

$$\sum_{i=1}^{s} \mu_1 (X_i) \,.\, Y_i = 0.$$

Come esempio, ripresa la prima serie ciclica del numero precedente, si avrà:

$$0{,}1295 \times 609 - 0{,}8495 \times 47 - 0{,}0652 \times 943 + 1{,}1951 \times 2593 + \\ + 0{,}6214 \times 30 - 0{,}3457 \times 2549 - 0{,}6856 \times 3229 = 0$$

ed anche:

$$0{,}1295\,(3229 + 47) - 0{,}8495\,(609 + 943) - 0{,}0652\,(47 + 2593) + \\ + 1{,}1951\,(943 + 30) + 0{,}6214\,(2593 + 2549) - 0{,}3457\,(30 + 3229) - \\ - 0{,}6856\,(2549 + 609) = 0.$$

Relazione fra i momenti primi e relazione fra i momenti secondi di una serie ciclica in due punti opposti del ciclo

34. — *a*) Se il punto O è interno ad uno degli intervalli $X'_t \ldots X'_{t+1}$ determinati dai punti opposti a quelle modalità di X per cui non è nullo il corrispondente valore di Y, sappiamo che per ogni X di scostamento x da O contenuto nello stesso intervallo è

$$\mu_1 (X) = \mu_1 (O) - \mu_0 x$$

Siano $X_h \ldots X_l$ le modalità di cui si incontrano i punti opposti $X'_h \ldots X'_l$: per andare da O al suo opposto O' nel senso positivo; ed $X_m \ldots X_p$ le modalità analoghe, andando da O ad O' in senso negativo. Allora il momento primo in un punto X di un intorno sufficientemente ristretto di O' è dato (*Vedi* n. 17) da

$$\mu_1 (X) = \mu_1 (O) - \mu_0 x + s (Y_h + \ldots + Y_l)$$

e, in particolare, sarà in O', cioè per $X = \frac{s}{2}$

$$\mu_1 (O') = \mu_1 (O) - \mu_0 \frac{s}{2} + s (Y_h + \ldots + Y_l)$$

$$= \mu_1 (O) + \frac{s}{2} (- Y_h \ldots - Y_l - Y_m \ldots - Y_p +$$

$$+ 2 Y_h + \ldots + 2 Y_l)$$

$$\mu_1 (O') = \mu_1 (O) + \frac{s}{2} (Y_h + \ldots Y_l - Y_m \ldots - Y_p) \tag{1}$$

Nel caso che O sia una media ordinaria si ha $\mu_1 (O) = 0$, onde :

$$\mu_1 (O') = \frac{s}{2} (Y_h + \ldots + Y_l - Y_m \ldots - Y_p)$$

b) Nelle stesse ipotesi di prima sarà per ogni X di $X'_t \ldots X'_{t+1}$
$\mu_2 (X) = \mu_0 x^2 - 2 \mu_1 (O) x + \mu_2 (O)$

e quindi per ogni X di un intorno sufficientemente ristretto di O' sarà per la (4) del numero 22:

$$\mu_2 (X) = \mu_0 x^2 - 2 \{ \mu_1 (O) + s (Y_h + \ldots + Y_l) \} x +$$
$$+ \{ \mu_2 (O) + s^2 (Y_h + \ldots + Y_l) + 2 s (x_h Y_h + \ldots x_l Y_l) \}$$

da cui si dedurrà in O', cioè per $x = \frac{s}{2}$

$$\mu_2 (O') = \mu_2 (O) + \frac{s^2}{4} (Y_h + \ldots + Y_l + Y_m + \ldots + Y_p) -$$
$$- s (x_h Y_h + \ldots + x_l Y_l + x_m Y_m + \ldots + x_p Y_p) +$$
$$+ s (2 x_h Y_h + \ldots + 2 x_l Y_l)$$

e infine

$$(2) \quad \mu_2 (O') = \mu_2 (O) + s (x_h Y_h + \ldots + x_l Y_l - x_m Y_m - x_p Y_p) + \\ + \frac{s^2}{4} \mu_o$$

Relazione fra le somme degli scostamenti assoluti di una serie ciclica da due punti opposti del ciclo.

35. — Si tratta di paragonare il comportamento della funzione $\Theta(X)$ (Cfr. n. 27) in due punti opposti del ciclo; e converrà distinguere il caso in cui il numero s delle modalità del ciclo sia pari, $s = 2k$, da quello in cui sia dispari, $s = 2k + 1$.

a) Quando sia $s = 2k$ (Cfr. n. 28) e si supponga che il punto O sia non esterno all'intervallo $X_t \ldots X_{t+1}$ il punto diametralmente opposto, O', sarà non esterno all'intervallo $X'_t \ldots X'_{t+1}$ che coincide con l'intervallo $X_{t+k} \ldots X_{t+k+1}$.

Ora, rispetto all'origine X_t, sia

$$(1) \qquad \Theta(x) = p_1 x + q_1$$

l'espressione della funzione somma degli scostamenti assoluti in $X_t \ldots X_{t+1}$; per quanto sappiamo circa il procedimento di passaggio dalla espressione di $\Theta(X)$ per un intervallo $X_t \ldots X_{t+1}$ a quella per l'intervallo successivo, è facile vedere che nell'intervallo

$$X'_t \ldots X'_{t+1} = X_{t+k} \ldots X_{t+k+1}$$

l'espressione di $\Theta(x)$ sarà

$$(2) \quad \Theta(x) = - p_1 x + \Big[q_1 + k p_1 + 2 \{ (k-1) Y_{t+1} + (k-2) Y_{t+2} + \\ + \ldots + 1 . Y_{t+k+1} \} - 2 \{ (k-1) Y_{t+k+1} + \\ + (k-2) Y_{t+k+2} + \ldots + 1 . Y_{t+2k} \} \Big]$$

dove è, come sempre, da intendersi che sia $Y_r = Y_t$ se $r \equiv t \pmod{s}$.

Se si intende che x sia lo scostamento di O rispetto ad X_t, sarà pure x lo scostamento di O' rispetto ad $X'_t = X_{t+k}$ cosicchè le (1) e (2), per tale significato di x forniscono

$$\Theta(O') = \Theta(O) - 2\,p_1\,x + \Big[k\,p_1 + 2\left\{(k-1)\,Y_{t+1} + \ldots + \right.$$
$$\left. + 1.\,Y_{t+k-1}\right\} - 2\left\{(k-1)\,Y_{t+k+1} + \ldots + 1.Y_{t+2k}\right\}\Big]$$

che è la relazione cercata.

b) Sia, in secondo luogo, $s = 2k + 1$ (Cfr. n. 29). Considerati allora i successivi intervalli $X_t \ldots X'_{t+k+1}$, $X'_{t+k+1} \ldots X_{k+1}$, $X_{t+t} \ldots X'_{t+k+2}$, ecc., si avrà, per quanto si disse circa l'espressione di $\Theta(X)$ nel passaggio da uno di tali intervalli al successivo, che se, rispetto all'origine X_t, è in $X_t \ldots X'_{t+k+1}$

(1) $$\Theta(x) = t_1\,x + u_1$$

sarà nell'intervallo successivo $X'_{t+k+1} \ldots X_{t+1}$ e rispetto all'origine X_{t+1}:

(2) $$\Theta(x) = (t_1 - 2\,Y_{t+k+1})\,x + (u_1 + t_1 - Y_{t+k+1})\,;$$

mentre nell'intervallo opposto al primo di questi, cioè in $X' \ldots X_{t+k+1}$ sarà rispetto all'origine X_{t+k+1}:

(3) $$\Theta(x) = -t_1\,x + (u_1 + t_1 - Y_{t+k+1} - Y_{t+k+2} \ldots - Y_{t+2k+1})$$

e nel successivo X_{t+k+1} sarà, sempre rispetto alla medesima origine X_{t+k+1}:

(4) $$\Theta(x) = (-t_1 + 2\,Y_{t+k+1})\,x +$$
$$+ (u_1 + t_1 - Y_{t+k+1} - Y_{t+k+2} - \ldots Y_{t+2k+1})$$

Ora, se O è non esterno all'intervallo $X_t \ldots X'_{t+k+1}$ ed ha lo scostamento x da X_t il suo opposto O' sarà non esterno ad $X'_t \ldots X_{t+k+1}$ ed avrà da X_{t+k+1} lo scostamento $x - 1$, cosicchè la (1), per il detto valore di x, e la (3) per sostituzione di $x - 1$ ad x, cioè la

(3') $$\Theta(x-1) = -t_1\,x + (u_1 + 2t_1 - Y_{t+k+1} \ldots - Y_{t+2k+1})$$

esprimeranno la somma degli scostamenti assoluti nei due punti O ed O'.

E, similmente, se O non è esterno all'intervallo $X'_{t+k+1} \ldots X_{t+1}$ ed ha lo scostamento x da X_{t+1}, il suo opposto O' sarà non esterno all'intervallo $X_{t+k+1} \ldots X'_{t+1}$, ed avrà rispetto all'origine X_{t+k+1} lo

scostamento $x+1$, onde la (2) per il detto valore di x e, la (4) per sostituzione di $x+1$ ad x, cioè la

$$(4') \qquad \Theta(x+1) = (-t_1 + 2Y_{t+k+1})\,x + \\ + (u_1 + Y_{t+k+1} - Y_{t+k+2} \dots - Y_{t+2k+1})$$

esprimeranno la somma degli scostamenti assoluti nei due punti O ed O'.

Indici di mutabilità di una serie ciclica

36. — Abbiamo trovato che una serie ciclica può ammettere una o più medie aritmetiche, e una o più mediane, nei diversi significati considerati; si presenta ora il problema di vedere se e come possano per essa definirsi degli indici di mutabilità basati sulla conoscenza di tali medie e mediane, e che siano analoghi a quegli indici di mutabilità che per una seriazione o per una serie rettilinea si fanno dipendere dalla sua media aritmetica o dalla sua mediana. Oltre a tali indici, si potranno, naturalmente, considerare anche quelli che si definiscono indipendentemente dalle modalità medie.

Scostamento semplice medio dalle medie aritmetiche

37. — Dato che una serie ciclica ammette generalmente più medie aritmetiche, e precisamente una media aritmetica, ordinaria, limite o fittizia in corrispondenza a ciascuno degli intervalli in cui il ciclo è diviso dai punti opposti alle modalità di X per cui non sia nullo il rispettivo valore di Y, (1) si potrà ovviamente definire lo scostamento semplice medio della data serie rispetto a ciascuna di quelle medie. Si avranno, in tal modo, altrettanti indici di mutabilità della serie, quante ne sono le medie, ma è chiaro che nessuno di essi avrà, almeno *a priori*, titolo di preferenza rispetto agli altri per indicare il comportamento della serie in quanto si riferisce alla sua mutabilità.

Si può, tuttavia, pensare alla formazione di un unico indice di variabilità in cui concorrano le diverse medie; e in quest'ordine

(1) Oltre le eventuali medie *sui generis*.

di idee si presenta molto naturale e fondata questa definizione: dal momento che in corrispondenza a ciascuno degli accennati intervalli esiste una media aritmetica (ordinaria, limite o fittizia) *si misuri lo scostamento assoluto che ogni modalità di X ha dalla media corrispondente all'intervallo in cui essa modalità cade, e la media di tali scostamenti assoluti, ponderati mediante i corrispondenti valori di* Y, *costituirà un indice di mutabilità della serie che si potrà dire «scostamento semplice medio della serie ciclica data dalle medie aritmetiche» e che indicheremo provvisoriamente con* ${}^1\eta$.

Per precisare il significato di questa definizione conviene distinguere il caso in cui il numero s delle modalità di X sia pari da quello in cui sia dispari, intendendo peraltro, come si è sempre fatto, che le modalità siano equispaziate.

Se s è dispari, ciascuna modalità X_i di X cade, per necessità, internamente ad uno degli intervalli anzidetti e quindi non vi è dubbio circa la media da cui si deve misurare lo scostamento di X_i.

Se s è pari, e cioè $s = 2k$, la modalità X_i coincide con X'_{i+k} ossia è opposta alla modalità X_{i+k}; allora se $Y_{i+k} = 0$, il punto X'_{i+k}, cioè X_i cadrà ancora *internamente* a uno dei soliti intervalli e il suo scostamento assoluto dovrà essere misurato dalla media che compete a questo intervallo; se, invece, $Y_{i+k} \neq 0$, il punto X_{i+k} cioè X_i *separerà* due di quegli intervalli fra loro contigui, di modo che vi sarà ambiguità circa la media da cui dovrà misurarsi lo scostamento di X_i; *converremo, in questa eventualità, di assumere in valore assoluto la media degli scostamenti algebrici che* X_i *ha da ciascuna delle due medie aritmetiche competenti agli intervalli separati dalla* X_i *stessa.*

Ciò posto, l'enunciata definizione fornisce come valore di ${}^1\eta$:

$$(1) \qquad {}^1\eta = \frac{\sum_{i=1}^{s} |\overline{X}_i X_i| \, Y_i}{\sum_{i=1}^{s} Y_i}$$

essendo $|\overline{X}_i X_i|$ lo scostamento assoluto della modalità X_i dalla media aritmetica $\overline{X}_i$ che compete allo stesso intervallo o, eventualmente, la semisomma in valore assoluto degli scostamenti algebrici dalle due medie aritmetiche che competono agli intervalli limitrofi.

Se dunque s è dispari si avrà, in valore algebrico, ricordando quanto si disse al numero 16,

(2) $(X_i \overline{X}_i)$ = scostamento dall'origine X_i della media $\overline{X}_i$ competente all'intervallo in cui cade X_i, =

$$= \frac{\mu_1(X_i)}{\mu_0} = \frac{\sum_{j=1}^{s} x_{ij} Y_j}{\sum_{j=1}^{s} Y_j},$$

essendo x_{ij} lo scostamento algebrico di X_j dalla modalità X_i presa come origine.

Altrettanto si potrà dire se s è pari ed X_i si oppone ad una modalità X_{i+k} per cui sia $Y_{i+k} = 0$, poichè in tal caso X_i cadrà internamente ad uno dei soliti intervalli e si potrà ivi assumere come origine, per la determinazione della media stessa, X_i.

Se, infine, s è pari ed X_i si oppone ad una modalità X_{i+k} per cui sia $Y_{i+k} \neq 0$ sarà, per la convenzione testè fatta,

(3) $(\overline{X}_i X_i)$ = semisomma degli scostamenti da X_i delle medie competenti agli intervalli limitrofi ad X_i =

$$= \frac{1}{2}\left(\frac{lims\ \mu_1(X_i)}{\mu_0} + \frac{limd\ \mu_1(X_i)}{\mu_0}\right)$$

e poichè sappiamo già (n. 16) che, essendo X_i punto di discontinuità del momento primo, risulta

$$\frac{1}{2}\left\{lims\ \mu_1(X_i) + limd\ \mu_1(X_i)\right\} = \mu_1(X_i),$$

così potremo, anche in questo caso, scrivere

$$(2) \qquad (X_i \overline{X}_i) = \frac{\mu_1(X_i)}{\mu_0} = \frac{\sum_{j=1}^{s} x_{ij} Y_j}{\sum_{j=1}^{s} Y_j},$$

come nella eventualità di s dispari.

Perciò sarà sempre

$$(1) \quad {}^1\eta = \frac{\sum_{i=1}^{s} \left|X_i \overline{X}_i\right| Y_i}{\sum_{i=1}^{s} Y_i} = \frac{\sum_{i=1}^{s} \dfrac{\left|\sum_{j=1}^{s} x_{ij} Y_j\right|}{\sum_{j=1}^{s} Y_j} Y_i}{\sum_{i=1}^{s} Y_i} = \frac{1}{\left(\sum_{i=1}^{s} Y_i\right)^2} \sum_{i=1}^{s} \left|\sum_{j=1}^{s} x_{ij} Y_j\right| Y_i$$

38. — Le considerazioni fatte nel numero precedente valgono comunque sia il carattere quantitativo Y, purchè a valori non negativi. Ma supponiamo, ora, che Y_i sia una frequenza, cioè il numero dei casi in cui si presenta, nella data serie ciclica, la modalità X_i È allora $\mu_o = \sum_{i=1}^{s} Y_j = n$, numero totale dei casi costituenti la serie ciclica, onde

$$(X_i \overline{X}_i) = \frac{\sum_{j=1}^{s} x_{ij} Y_j}{\sum Y_j} = S_{R,i}$$

essendo $S_{R,i}$ ciò che fu già altrove (1) definito come *scostamento con ripetizione del carattere* X, in funzione del quale è costruita la serie, *nel caso* 1°, o, ciò che è lo stesso, *scostamento con ripetizione della modalità* X$_i$ *di* X. Perciò *quello che si dice* in una serie ciclica di frequenze, *scostamento con ripetizione della modalità* X_i, *non è altro*, secondo quanto abbiamo visto dalle precedenti considerazioni, *che lo scostamento della modalità* X_i *dalla media competente all'intervallo in cui* X_i *giace, o la semisomma degli scostamenti di* X_i *dalle medie competenti agli intervalli limitrofi, se* X_i *separa due di tali intervalli.*

Si ha, inoltre, che

$$(1'') \qquad \frac{1}{n} \sum_{i=1}^{s} | S_{R,j} | \; Y_i = {}^1S_R$$

essendo ${}^1S_R$ ciò che venne già altrove (2) definito come *scostamento semplice medio con ripetizione della serie data*, e si sa anche, d'altra parte, che:

$$\frac{1}{n} \sum_{i=1}^{s} | S_{R,i} | \; Y_i = \frac{\sum_{i=1}^{s} | X_i \overline{X}_i | \; Y_i}{\sum_{i=1}^{s} Y_i} = {}^1\eta .$$

Pertanto, *quello che si dice scostamento semplice medio con ripetizione di una serie ciclica coincide con quello che abbiamo sopra definito come « scostamento semplice medio della serie ciclica dalle medie aritmetiche »*, cosicchè potremmo d'ora in poi abbandonare il simbolo ${}^1\eta$ e mantenere quello già in uso ${}^1S_R$, tanto più che a questo fa riscontro un indice analogo ${}^1S$ che è lo *scostamento semplice medio* (*senza ripetizione*) *della data serie;* conserveremo, tuttavia, anche

(1) C. GINI. *Di una estensione del concetto di scostamento medio*, ecc. già citato.

(2) Luogo testè citato.

la notazione ${}^{\mathrm{I}}\eta$, per ragione di simmetria con altri simboli che adopreremo.

Inoltre si ricordi che nella rappresentazione geometrica da noi data per il momento primo di una serie ciclica, (Cfr. n. 19), essendosi assunto $\Sigma\, Y_i$ come unità di misura sull'asse delle Y, l'ordinata di ciascun punto della spezzata rappresentativa di $\mu_{\mathrm{I}}\,(X)$ segna la distanza di quel punto dalla media che giace nello stesso intervallo (salvo la solita modificazione se il punto stesso separa due di quegli intervalli fra loro contigui). Ne segue che *geometricamente l'indice di mutabilità* ${}^{\mathrm{I}}\mathrm{S}_{\mathrm{R}}$ *è rappresentato dalla media ponderata delle ordinate assolute di* μ_{I} *nei punti* X_i, *attribuendosi alla ordinata in* X_i *il peso* Y_i.

Osservazione 1ª. — Se di una seriazione o di una serie rettilinea consideriamo lo scostamento semplice medio dalla media aritmetica A, cioè:

$$(4) \qquad {}^{\mathrm{I}}\eta_A = \frac{\sum\limits_{i=1}^{s} |AX_i|\, Y_i}{\sum\limits_{i=1}^{s} Y_i}$$

e paragoniamo questa espressione con quella (1) di ${}^{\mathrm{I}}\eta$ nel caso di una serie ciclica, vediamo che le (1) e (4) formalmente coincidono, salvo in ciò: *che nella* (4) *gli scostamenti delle varie modalità di* X *sono tutti misurati rispetto all'unica media* A *esistente nella seriazione o nella serie rettilinea; mentre nella* (1) *lo scostamento di ogni modalità* X_i *è misurato rispetto a quella media aritmetica che compete all'intervallo in cui giace* X_i, *se* s *è dispari, e rispetto alla semisomma delle medie competenti agli intervalli adiacenti ad* X_i, *se* s *è pari.*

Osservazione 2ª. — Nella formazione dell'indice di mutabilità espresso dalle (1) e (1'') non entrano in calcolo quelle modalità di X per le quali sia nullo il corrispondente valore di Y, e quindi neanche quelle medie aritmetiche che competono agli intervalli in cui giacciono quelle modalità. In particolare *se tutte le modalità per cui non è nullo il valore di* Y *giacciono internamente ad un emiciclo, nel calcolo dell'indice* (1) *ed* (1'') *entrano soltanto queste modalità e quell'unica media giacente nello stesso emiciclo, che già venne particolarmente segnalata nell'osservazione quarta del numero* 18. *In questo caso la* (1) *si identifica perfettamente con ciò che per una seriazione o per una serie rettilinea è lo scostamento semplice medio dalla media; cosicchè per una serie ciclica così fatta l'indice* (1) *si può ben dire*

scostamento semplice medio dalla media aritmetica di cui alla osservazione precitata.

Osservazione 3ª. — Il fatto che lo scostamento semplice medio con ripetizione di una serie ciclica è stato definito dapprima indipendentemente dal concetto di media aritmetica, ma se ne è poi trovata una semplice relazione con quelle che sono le medie aritmetiche, effettive e fittizie, della serie stessa, costituisce un nuovo argomento a favore dell'opportunità di introdurre tali medie, *anche quando esse non coincidono* (e questo è appunto il caso più generale) *con nessuna delle modalità del carattere* X *in funzione del quale è data la serie ciclica* (Cfr. n. 13). Si vede inoltre che anche le medie fittizie concorrono alla formazione di ${}^{1}\eta = {}^{1}S_R$.

Osservazione 4ª. — In una serie ciclica lo scostamento semplice medio dalle medie aritmetiche ${}^{1}\eta = {}^{1}S_R$ è generalmente minore di ciascuno degli scostamenti semplici medî che potrebbero essere calcolati rispetto alle singole medie (Cfr. n. 35), perchè gli scostamenti delle varie modalità dalle medie che competono agli intervalli in cui cadono esse modalità sono generalmente minori degli scostamenti delle stesse modalità, misurati tutti da un'unica media. Tuttavia, se, come nella osservazione 2ª, tutte le modalità alle quali corrispondono valori non nulli di Y cadono in uno stesso emiciclo, lo scostamento ${}^{1}\eta = {}^{1}S_R$ coincide con lo scostamento semplice medio della data serie rispetto a quella media (ordinaria) che compete all'emiciclo in cui sono quelle modalità. Questa condizione di minimo a cui soddisfa ${}^{1}\eta = {}^{1}S_R$ corrobora la preferenza accordata a ${}^{1}S_R$ come indice di mutabilità della serie, piuttosto che agli scostamenti semplici medî *dalle singole medie* o ad una media di questi scostamenti medî.

Osservazione 5ª. — Al n. 16 sono state denominate « inaccettabili » le medie aritmetiche limite e fittizie, poichè in esse il momento primo non si annulla: nondimeno esse concorrono a formare l'indice ${}^{1}\eta = {}^{1}S_R$, ed anche l'indice ${}^{2}\eta = {}^{2}S_R$ di cui al numero seguente. Non vi concorrono invece, benchè accettabili, le medie *sui generis*.

Scostamento quadratico medio dalle medie aritmetiche

39. — In analogia alla definizione di scostamento semplice medio di una serie ciclica dalle medie aritmetiche, *la media quadratica ponderata degli scostamenti che ogni modalità di* X *ha dalla media*

corrispondente all'intervallo in cui essa cade, quando si assuma come peso, per ogni modalità, il corrispondente valore di Y, costituirà un indice di mutabilità della serie che si dirà « scostamento quadratico medio della data serie ciclica dalle medie aritmetiche » e che si indicherà provvisoriamente con $^2\eta$.

Sarà dunque:

$$(5) \qquad {}^2\eta = \left(\frac{\sum\limits_{i=1}^{s} (\overline{X}_i\, X_i)^2\, Y_i}{\sum\limits_{i=1}^{s} Y_i} \right)^{\frac{1}{2}}$$

indicando, come già si fece, con $(\overline{X}_i\, X_i)$ lo scostamento della modalità X_i dalla media aritmetica $\overline{X}_i$ che compete allo stesso intervallo in cui cade X_i se tale modalità è interna a uno dei soliti intervalli determinati dai punti opposti alle modalità per cui non è nullo il corrispondente valore di Y; e intendendo, invece, che $(\overline{X}_i\, X_i)$ sia la semisomma degli scostamenti di X_i dalle medie che competono ai due intervalli limitrofi, se X_i separa questi due intervalli.

Sarà quindi, in tutti i casi:

$$(5') \qquad {}^2\eta = \left(\frac{\sum\limits_{i=1}^{s} \left(\frac{\sum\limits_{j=1}^{s} x_{ij}\, Y_j}{\sum\limits_{j=1}^{s} Y_j} \right)^2 Y_i}{\sum\limits_{i=1}^{s} Y_i} \right)^{\frac{1}{2}} = \frac{1}{\sum\limits_{i=1}^{s} Y_i} \left(\frac{\sum\limits_{i=1}^{s} \left(\sum\limits_{j=1}^{s} x_{ij}\, Y_j \right)^2 Y_i}{\sum\limits_{i=1}^{s} Y_i} \right)^{\frac{1}{2}}$$

dove x_{ij} denota lo scostamento algebrico di X_j dalla modalità X_i presa come origine.

Se Y_i è il numero dei casi in cui si presenta, nella data serie ciclica, la modalità X_i, dimodochè

$$\mu_0 = \sum_{i=1}^{s} Y_i = n \text{ (numero totale dei casi)},$$

allora sappiamo già che

$$\frac{\sum\limits_{j=1}^{s} x_{ij}\, Y_j}{\sum\limits_{j=1}^{s} Y_j} = S_{R,i}$$

(Cfr. numero precedente) dimodochè $^2\eta$ *è la media quadratica degli scostamenti con ripetizione delle modalità* X_i *di* X; d'altronde tale media ha già ricevuto la denominazione di « *scostamento quadratico medio*

con ripetizione della serie » (1) *e viene designata col simbolo* ${}^2S_R$, cosicchè quello che si dice scostamento quadratico medio con ripetizione di una serie ciclica coincide con quello che abbiamo testè definito come « scostamento quadratico medio della serie ciclica dalle medie aritmetiche », e si potrà scrivere ${}^2\eta = {}^2S_R$.

Non ci soffermiamo a dare la facile interpretazione geometrica di ${}^2S_R$.

Insistiamo, piuttosto, nel rilevare che la posta definizione di ${}^2\eta$, per una serie ciclica, costituisce una ben naturale estensione della solita definizione di scostamento quadratico medio di una seriazione o di una serie rettilinea dalla media aritmetica, con la sola differenza che per una serie ciclica gli scostamenti delle varie modalità sono misurati dalle diverse medie aritmetiche e per una seriazione o per una serie rettilinea dall'unica media aritmetica ivi esistente.

Notiamo pure che *se tutte le modalità per cui non è nullo il valore di* Y *giacciono in uno stesso emiciclo, nel calcolo di* ${}^2\eta = {}^2S_R$ *entrano soltanto queste modalità e quell'unica media giacente nello stesso emiciclo, di cui fu fatta parola nella osservazione* 4ª *del numero* 18; *ed in questo caso l'indice* ${}^2\eta = {}^2S_R$ *si identifica perfettamente con ciò che per una seriazione o per una serie rettilinea è lo scostamento quadratico medio dalla media.*

Osservazione. — Se si calcolassero gli scostamenti quadratici medî della serie ciclica data dalle sue singole medie aritmetiche, si troverebbero valori generalmente superiori a quello di ${}^2\eta = {}^2S_R$: il che costituisce una ragione di preferenza di questo indice a paragone di quegli scostamenti quadratici medî. Ciò non significa, tuttavia, che non possa, talora, interessare anche la considerazione degli scostamenti quadratici medî dalle singole medie aritmetiche.

Se, dal numero 22, si riprende l'espressione del momento secondo in X, valida nell'intervallo $X'_t\, X'_{t+1}$, e cioè:

$$\mu_2\,(X) = \mu_0\, x^2 - 2\,\mu_1\,(O)\,.\,x + \mu_2\,(O),$$

essendo x lo scostamento di X da O, e si suppone che l'origine O coincida con la media aritmetica A relativa a quell'intervallo, si avrà $\mu_1\,(A) = 0$ e quindi

$$\mu_2\,(X) = \mu_2\,(A) + \mu_0\, x^2$$

$$\frac{\mu_2\,(X)}{\mu_0} = \frac{\mu_2\,(A)}{\eta_0} + x^2$$

$${}^2\eta_X{}^2 = {}^2\eta_A{}^2 + x^2$$

(1) C. Gini. *Di una estensione del concetto di scostamento medio*, ecc. già citato.

che è la nota relazione pitagorica fra lo scostamento quadratico medio ${}^2\eta_A$ della serie dalla media aritmetica A, lo scostamento quadratico medio ${}^2\eta_X$ da un'altra modalità X, e lo scostamento x di X da A.

Una relazione di questa forma vale dunque, per una serie ciclica, internamente a ciascuno degli intervalli come $X'_t \dots X'_{t+1}$; mentre per una seriazione e quindi anche per una serie rettilinea la relazione è, come si sa, unica.

Esempi

40. — Facciamo applicazione dei concetti esposti ad alcuni esempi.

Esempio 13°. — Per la serie ciclica degli esempi 4°, 5°, 9°, 12°, e cioè:

Modalità di X:	L	Ma	Me	G	V	S	D
Valori di Y:	5	2	1	6	3	0	5

di cui si sa (*Vedi* es. 5°) che le medie (ordinarie, limite e fittizie) hanno da queste modalità i rispettivi scostamenti:

$$x_{IV} = \frac{8}{22}, \; x_V = \frac{7}{22}, \left(-\frac{15}{22}\right) (1), \; x_{VI} = -\frac{2}{22},$$

$$x_I = \frac{11}{22}, \; x_{II} = \frac{3}{22}, \; x_{III} = -\frac{12}{22}$$

sarà:

$${}^1S_R = {}^1\eta = \frac{1}{22}\left(5 \cdot \frac{8}{22} + 2 \cdot \frac{7}{22} + 1 \cdot \frac{15}{22} + 6 \cdot \frac{2}{22} + \right.$$

$$\left. + 3 \cdot \frac{11}{22} + 5 \, \frac{12}{22}\right) = \frac{174}{484}.$$

(1) La media avente da Me lo scostamento $-\frac{15}{22}$ è, come sappiamo, quella stessa X_V che ha da Ma lo scostamento $\frac{7}{22}$.

Questo valore è minore degli scostamenti semplici medî della serie rispetto alle singole medie, scostamenti che potranno indicarsi con ${}^{1}\eta_{x_{I}}$, ${}^{1}\eta_{x_{II}}, \ldots$ ecc. Si trova, infatti,

$$
{}^{1}\eta_{x_{I}} = \frac{880}{484};\ {}^{1}\eta_{x_{II}} = \frac{845}{484};\ {}^{1}\eta_{x_{III}} = \frac{838}{484};\ {}^{1}\eta_{x_{IV}} = \frac{824}{484};
$$

$$
{}^{1}\eta_{x_{V}} = \frac{828}{484};\ {}^{1}\eta_{x_{VI}} = \frac{848}{484}.
$$

Per la stessa serie si trova

$$
{}^{2}S_{R} = {}^{2}\eta = \left\{\frac{1}{22}\,\frac{5.64 + 2.49 + 1.225 + 36.4 + 3.121 + 5.144}{484}\right\}^{\frac{1}{2}} =
$$

$$
= \frac{1}{22}\sqrt{\frac{1870}{22}}.
$$

Gli scostamenti quadratici medî rispetto alle singole medie superano tutti maggiori di ${}^{2}S_{R}$; ed invero:

$$
{}^{2}\eta_{x_{I}} = \frac{1}{22}\sqrt{\frac{42350}{22}};\ {}^{2}\eta_{x_{II}} = \frac{1}{22}\sqrt{\frac{37383}{22}};\ {}^{2}\eta_{x_{III}} = \frac{1}{22}\sqrt{\frac{39116}{22}};
$$

$$
{}^{2}\eta_{x_{IV}} = \frac{1}{22}\sqrt{\frac{43120}{22}};\ {}^{2}\eta_{x_{V}} = \frac{1}{22}\sqrt{\frac{36190}{22}};\ {}^{2}\eta_{x_{VI}} = \frac{1}{22}\sqrt{\frac{49280}{22}}.
$$

Esempio 14°. — Per la serie ciclica dell'esempio 7°, e cioè:

Modalità di X:	L	Ma	Me	G	V	S	D
Valori di Y:	5	2	0	7	0	0	0

si sa che le medie hanno dalle singole modalità i rispettivi scostamenti:

$$
\left(\frac{23}{14}\right),\ x_{III} = \frac{9}{14},\ \left(-\frac{5}{14}\right),\ \left(-\frac{19}{14}\right),\ x_{I} = \frac{2}{14},\ x_{II} = \frac{2}{14},\ \left(-\frac{12}{14}\right)
$$

e perciò

$$
{}^{1}S_{R} = {}^{1}\eta = \frac{1}{14}\left(5\cdot\frac{23}{14} + 2\cdot\frac{9}{14} + 7\cdot\frac{19}{14}\right) = \frac{296}{196}
$$

mentre

$$
{}^{1}\eta_{x_{I}} = \frac{400}{196};\ {}^{1}\eta_{x_{II}} = \frac{420}{196};\ {}^{1}\eta_{x_{III}} = \frac{296}{196}.
$$

Si trova pure:

$$^2S_R = {}^2\eta = \left\{\frac{1}{14}\,\frac{5.529+2.81+7.361}{196}\right\}^{\frac{1}{2}} = \frac{1}{14}\sqrt{\frac{5334}{14}}$$

mentre

$$^2\eta_{x_{\mathrm{I}}} = \frac{1}{14}\sqrt{\frac{12880}{14}}\,,\quad {}^2\eta_{x_{\mathrm{II}}} = \frac{1}{14}\sqrt{\frac{13664}{14}}\,,\quad {}^2\eta_{x_{\mathrm{III}}} = \frac{1}{14}\sqrt{\frac{5234}{14}}$$

Poichè le tre modalità per cui non è nullo il valore di Y giacciono in uno stesso emiciclo, e quella particolare media giacente in tale emiciclo è quella indicata da x_{III} si trova appunto:

$$^1S_R = {}^1\eta = {}^1\eta_{x_{\mathrm{III}}}\,;\quad {}^2S_R = {}^2\eta = {}^2\eta_{x_{\mathrm{III}}}\,,$$

e inoltre:

$$({}^1\eta_{x_{\mathrm{I}}}\,,\ {}^1\eta_{x_{\mathrm{II}}}) > {}^1S_R \quad ({}^2\eta_{x_{\mathrm{I}}}\,,\ {}^2\eta_{x_{\mathrm{II}}}) > {}^2S_R\,.$$

Scostamento semplice medio dalla mediana

41. — Come, per una seriazione o per una serie rettilinea, ha particolare interesse quell'indice di variabilità che è costituito dallo « scostamento semplice medio dalla mediana » in quanto lo scostamento semplice medio di tale seriazione o serie acquista il suo minimo valore se venga per l'appunto calcolato rispetto alla mediana, così si presenta ovviamente l'opportunità di considerare, anche per una serie ciclica, lo scostamento semplice medio $^1\eta_M$ rispetto alla mediana.

Ora, se la mediana è unica e si denota con M, sarà senz'altro, secondo le notazioni adottate al numero 27:

$$^1\eta_M = \frac{\sum\limits_{i=1}^{s} |M\ X_i|\,Y_i}{\sum\limits_{i=1}^{s} Y_i} = \frac{\Theta\ (M)}{\mu_0}\,.$$

Se invece la serie ammette diverse mediane in senso lato, sarà il caso di prescegliere fra queste le sole mediane in senso stretto, rispetto alle quali $\Theta\ (X)$ acquista uno stesso valore minimo assoluto, cosicchè il calcolo di $^1\eta_M$ rispetto ad una qualunque di tali mediane fornirà sempre un medesimo valore, che si potrà ancora dire scostamento semplice medio della data serie rispetto « alla mediana ».

Naturalmente, si potrebbero ben anche calcolare gli scostamenti semplici medi della data serie dalle altre mediane, ma tali scostamenti non avrebbero molto interesse nello studio della mutabilità della serie, non ottemperando a quella condizione di minimo (assoluto) che costituisce una ragione di preferenza per lo scostamento medio calcolato rispetto alla mediana. E si potrebbero altresì calcolare gli scostamenti quadratici medi rispetto alle varie mediane, o rispetto a quella o a quelle sole per cui tali scostamenti quadratici risultano minimi assoluti.

Se, infine, tutte le modalità di un intervallo si potessero considerare come mediane della serie data (cfr. n. 18), lo scostamento semplice medio rispetto a ciascuna di esse risulterebbe costante, per il modo stesso in cui le mediane si sono definite.

Esempio 15°. — La serie degli esempi 6°, 8°, 11°, ammette una sola mediana, costituita dalla modalità W per cui è $\Theta(W) = 50$; perciò ${}^1\eta_M = \frac{50}{30}$; inoltre si trova:

$$^2\eta_M = \left(\frac{3.2^2 + 5.3^2 + 0.4^2 + 4.3^2 + 2.2^2 + 7.1 + 3.0 + 5.1}{30}\right)^{\frac{1}{2}} =$$
$$= \sqrt{\frac{113}{30}}.$$

La serie degli esempi 9° e 12° ammette due mediane in senso stretto, D e L, ed è $\Theta(D) = 36$; $\Theta(L) = 36$; perciò ${}^1\eta_M = \frac{36}{22} = \frac{18}{11}$. Inoltre si trova, rispetto alla mediana D:

$$^2\eta'_{M=D} = \left(\frac{\mu_2(D)}{\mu_0}\right)^{\frac{1}{2}} = \sqrt{\frac{88}{22}},$$

e rispetto alla mediana L:

$$^2\eta''_{M=L} = \left(\frac{\mu_2(L)}{\mu_0}\right)^{\frac{1}{2}} = \sqrt{\frac{92}{22}}.$$

Il fatto che i valori degli scostamenti semplici rispetto alle due mediane coincidono non porta, naturalmente, a concludere che coincidano anche gli scostamenti quadratici.

Differenza media semplice e quadratica di una serie ciclica

42. — Sono già stati altrove proposti (1), come indici di mutabilità di una serie ciclica, la differenza semplice media senza e con ripetizione Δ e Δ_R e la differenza quadratica media senza e con ripetizione ${}^2\Delta$ e ${}^2\Delta_R$. Possedendo la rappresentazione grafica di $\Theta(X)$ e di $\mu_2(X)$ è facile interpretare geometricamente gli indici Δ_R e ${}^2\Delta_R$. Difatti quella che si dice « differenza con ripetizione del carattere X nel caso i^o », cioè « nella modalità X_i » non è altro che $\dfrac{\Theta(X_i)}{\mu_o}$ mentre « la differenza quadratica media con ripetizione di X nella modalità X_i » sarà data da $\left(\dfrac{\mu_2(X_i)}{\mu_o}\right)^{\frac{1}{2}}$.

Gli indici di mutabilità Δ_R e ${}^2\Delta_R$ non saranno altro che la media aritmetica ponderata dei valori $\dfrac{\Theta(X_i)}{\mu_o}$ e la media quadratica ponderata dei valori $\left(\dfrac{\mu_2(X_i)}{\mu_o}\right)^{\frac{1}{2}}$ quando il peso sia, per ogni X_i, costituito dal corrispondente valore Y_i. Perciò:

$$\Delta_R = \frac{\sum\limits_{i=1}^{s} \dfrac{\Theta(X_i)}{\mu_o} Y_i}{\sum\limits_{i=1}^{s} Y_t} = \frac{1}{\mu_o^2} \sum_{i=1}^{s} \Theta(X_i) . Y_i$$

$${}^2\Delta_R = \left(\frac{\sum\limits_{i=1}^{s} \dfrac{\mu_2(X_i)}{\mu_o} Y_i}{\sum\limits_{i=1}^{s} Y_i} \right)^{\frac{1}{2}} = \left(\frac{1}{\mu_o^2} \sum_{i=1}^{s} \mu_2(X_i) . Y_i \right)^{\frac{1}{2}}$$

Poichè $\Theta(X_i)$ e $\mu_2(X_i)$ sono le ordinate in X_i delle rappresentazioni geometriche date per la somma degli scostamenti assoluti e per il momento secondo, si vede appunto che i due indici Δ_R e Δ^2_R non sono altro che medie opportunamente ponderate delle ordinate stesse.

(1) C. Gini. *Di una estensione del concetto di scostamento medio*, ecc. già citato.

IV. — SERIE SCONNESSE

Generalità

43. — Se per le serie rettilinee e per le serie cicliche il concetto di ordine a cui soddisfano le modalità del carattere qualitativo X, in funzione del quale esse sono costruite, implica la possibilità di far corrispondere a tali modalità dei numeri tali che a modalità equidiverse facciano riscontro numeri equidifferenti, per le serie sconnesse manca assolutamente la possibilità di stabilire un'analoga corrispondenza fra numeri e modalità del carattere qualitativo X da cui dipendono.

Difatti, non soltanto queste modalità non presentano un ordine naturale di successione, cosicchè assuntane convenzionalmente una come prima non resta in alcun modo determinata la seconda, e fissata convenzionalmente la seconda non ne consegue quale sia la terza, ecc., ma in generale non avrebbe neanche un preciso significato il dire che la diversità fra due modalità sia maggiore o minore della diversità fra due altre; onde, astraendo dalla possibilità di eseguire un tale confronto fra queste diversità, si ammette, nello schema teorico di una serie sconnessa, che la diversità fra due modalità qualunque di X sia costante, ossia che la serie sia necessariamente equispaziata. Segue anche che, mentre nelle serie ordinate la diversità della modalità A alla B del carattere (ordinato) X si considera opposta alla diversità da B ad A, per una serie sconnessa la diversità fra due modalità A e B del carattere (sconnesso) X non dipende dall'ordine in cui esse si enunciano. La costante diversità fra due modalità qualsiasi di X si potrà rappresentare con un numero positivo, costante, p. es. l'unità positiva, che diremo « scostamento » fra le modalità stesse.

Pertanto, se si volessero geometricamente rappresentare le modalità di un carattere X sconnesso si potrebbe, se tali modalità fossero in numero di 2, 3, 4 assumere rispettivamente gli estremi di un segmento arbitrario, i vertici di un triangolo equilatero, i vertici di un tetraedro regolare; per un maggior numero di modalità verrebbe a mancare una rappresentazione nello spazio ordinario, e si dovrebbe necessariamente far ricorso ad uno spazio di ordine superiore; e precisamente, se le modalità fossero in numero di s le loro

immagini potrebbero essere costituite dai vertici di un s-gono regolare in uno spazio ad s-1 dimensioni.

Se poi si volesse completare la rappresentazione, così che questa comprendesse anche le modalità del carattere quantitativo Y dipendente da X, modalità che supporremo sempre positive, allora bisognerebbe ad ogni vertice attribuire un peso uguale al valore di Y corrispondente a quella modalità di X che esso rappresenta.

Per la designazione pratica delle s modalità di X e delle corrispondenti di Y si potranno assumere le notazioni

$$X_1 \, X_2 \ldots X_s$$
$$Y_1 \, Y_2 \ldots Y_s$$

senza, peraltro, intendere che il succedersi degli indici 1, 2, ...s debba significare una particolare conformazione della serie data; in altri termini le varie modalità di X si potrebbero designare mediante speciali denominazioni e gli indici ordinali 1, 2, ...s avrebbero il solo compito di denotare l'ordine arbitrario in cui vengono enunciate quelle denominazioni.

Media aritmetica

44. — Ciò posto, proponiamoci di estendere anche alle serie sconnesse, se e come sia possibile, il concetto di media aritmetica.

È escluso, *a priori*, che ciò possa farsi in base alla proprietà $\Sigma\varepsilon = 0$, ossia che la media aritmetica possa definirsi come una modalità di X per cui sia nulla la somma algebrica degli scostamenti dalle altre modalità, moltiplicati, tali scostamenti, per i corrispondenti valori di Y; difatti l'annullarsi di tale somma, dato che i valori di Y sono, per ipotesi, tutti positivi, richiederebbe che alcuni scostamenti fra le varie modalità di X fossero positivi ed altri negativi, il che è contro la natura della serie sconnessa.

Bisogna anche escludere che una tale condizione, non verificabile per nessuna delle date modalità di X, possa eventualmente verificarsi per altre modalità di X che si potessero opportunamente definire, come accadeva per le serie rettilinee o cicliche in cui una media o le medie aritmetiche potevano identificarsi con delle modalità di conto.

Una tale estensione del concetto di modalità, possibile ed anzi opportuna per le serie ordinate, in cui è precisamente l'esistenza di

un ordine fra le varie modalità a fornire il mezzo di introdurre fra le effettive altre modalità fittizie di X, *non potrebbe in nessun modo essere fatta per una serie sconnessa*, in cui non si saprebbe quale significato attribuire alla locuzione « modalità A compresa fra due date modalità B e C », tanto più che la modalità stessa A dovrebbe anche avere la proprietà di essere equidifferente da tutte le altre modalità di X; e quindi ad uguale titolo, si dovrebbe poter dire simultaneamente compresa anche « fra » due qualunque di queste altre modalità. E se la rappresentazione geometrica, testè accennata, potesse far pensare alla possibilità di considerare anche altri punti oltre i vertici di quel certo s-gono, come rappresentativi di nuove modalità fittizie di X, oltre quelle effettive, basterebbe considerare che non esiste nel medesimo spazio ad n-1 dimensioni, nessun altro punto, avente da tutti quei vertici distanze uguali fra di loro ed eguali a quelle che già sussistono fra i vertici stessi, per eliminare nel modo più assoluto ogni dubbio circa la possibilità di introdurre nuove modalità del carattere sconnesso X; anche l'introduzione di una sola modalità oltre le date richiederebbe il passaggio ad uno spazio ad s dimensioni e non si avrebbe più la serie sconnessa data, ma una nuova serie sconnessa che si potrebbe, tutt'al più, riguardare come comprensiva della prima.

Scartata, pertanto, la possibilità che una definizione di media aritmetica basata sulla proprietà $\Sigma\varepsilon = 0$ sia suscettibile di un significato per una modalità qualsiasi del carattere X da cui dipende la data serie sconnessa, o sia suscettibile di acquistarlo in riferimento a nuove modalità fittizie di X, rimane a vedere se, assumendo come definizione di media un'altra proprietà caratteristica, tale definizione possa, nel nostro caso, acquistare un significato concreto.

È ovvio che se assumiamo come definizione di media la $\Sigma\varepsilon^2 = min.$, se cioè *diciamo media aritmetica della data serie sconnessa la modalità od una modalità X_i di X per cui risulti*

$$\sum_{j=1}^{s} (X_i\, X_j)^2\, Y_j = min. \tag{1}$$

allora, poichè

$$(X_i\, X_j) = 1 \ (\text{se } i \neq j) \ \text{ ed } \ (X_i\, X_j) = 0 \ (\text{se } i = j)\ ,$$

la condizione posta si potrà scrivere:

$$\sum_{j=1}^{i-1} Y_i + \sum_{j=i+1}^{s} Y_j = min.\,, \tag{1'}$$

che equivale alla

$$(1'') \qquad Y_i = max.,$$

onde si concluderà che *tale definizione è verificata per quella modalità o per una qualunque delle modalità di* X *alle quali compete il massimo valore di* Y.

In questo senso una serie sconnessa ammette sempre una almeno ed eventualmente più medie aritmetiche.

Se la data è una serie di frequenze, la media aritmetica sarà costituita da una qualunque delle modalità di X per cui la frequenza sia massima.

Finalmente, in analogia al significato meccanico di media aritmetica di una seriazione o di una serie rettilinea, si potrebbe pensare a definire la media aritmetica di una serie sconnessa come il baricentro del sistema di punti rappresentativi delle varie modalità di X, quando ciascuno di tali punti abbia un peso rappresentato dal corrispondente valore di Y; ma, dato che tale baricentro non potrebbe coincidere con nessuna delle date modalità, si ricadrebbe daccapo nella necessità di dover pensare che punti distinti dai rappresentativi delle modalità, possano essere immagini di nuove modalità di X, ciò che non può verificarsi.

In conclusione, *il solo mezzo per estendere alle serie sconnesse il concetto di media aritmetica è quello di definire tale media in base alla proprietà* $\Sigma\varepsilon^2 = min.$

Moda e mediana di una serie sconnessa

45. — Nessuna particolarità offre, per una serie sconnessa, la definizione di moda, che continua ad essere come già per le seriazioni e per le serie ordinate, la modalità od una qualunque delle modalità, a cui corrisponde il massimo valore di Y, e, in particolare la massima frequenza se si tratti di una serie di frequenze.

Il concetto di mediana esige, invece, per essere esteso alle serie sconnesse, una considerazione analoga a quella che già si fece per le serie cicliche. Non esistendo una prima (ed un'ultima) modalità di X, non si può dare quella solita definizione di mediana che implica il succedersi delle modalità di X in un ordine naturale a partire da una di esse.

Se poi si volessero disporre le varie modalità in un certo ordine, si avrebbe una « mediana » in corrispondenza a ciascuno dei possibili $s!$ ordinamenti (ed anzi ciascuna delle date modalità potrebbe essere mediana di convenienti ordinamenti), ma nessuna delle mediane stesse potrebbe avere titolo di preferenza rispetto alle altre.

Tuttavia, se si riflette che una proprietà caratteristica della mediana è espressa dalla

$$(2) \qquad \sum_{j=1}^{s} |X_i X_i| \, Y_j = min.$$

e che il primo membro della (2) acquista un valore ben determinato per ciascuna modalità di X, allora si concluderà che *per una serie sconnessa si può definire come mediana quella modalità od una qualunque di quelle modalità* X_i *per cui sia verificata la* (2) *e che in tale senso la mediana di una serie sconnessa è costituita da quella o da una qualunque di quelle modalità di* X *per cui sia massimo il corrispondente valore di* Y.

Coincidenza di media aritmetica moda e mediana

46. — I numeri 44 e 45 ci hanno portato a concludere che per una serie sconnessa non soltanto i diversi concetti di media aritmetica, di moda e di mediana si possono opportunamente estendere, ma tutti conducono a definire una stessa o delle stesse modalità di X simultaneamente come media aritmetica, moda e mediana della serie data.

Un'altra osservazione deve essere fatta prima di lasciare questo argomento, e si riferisce alla circostanza che le definizioni (1) e (2) di media aritmetica e di mediana avrebbero un significato analitico ben diverso da quello a cui noi siamo stati costretti dalla necessità di ricercare la media e la mediana fra le sole modalità di X, se si potessero, in riferimento alla solita rappresentazione geometrica in uno spazio ad n-1 dimensioni, considerare anche gli altri punti del medesimo spazio, distinti dai vertici del solito s-gono.

Essendo $x_1\, x_2 \ldots x_{s-1}$ le coordinate di un punto generico X in uno spazio ad s-1 dimensioni, ed $x_{1j}, x_{2j}, \ldots x_{s-1,j}$ le coordinate

di X_j la condizione (1) scritta per un punto generico X (x_1, x_2, ... x_{s-1}) invece che pei un punto-modalità X_i, e cioè la

$$\varphi(x_1\, x_2 \ldots x_{s-1}) = \sum_{j=1}^{s} (XX_j)^2\, Y_j = min.$$

in cui

$$(X_i\, X_j)^2 = (x_{1,j} - x_1)^2 + \ldots + (x_{s-1,j} - x_{s-1})^2$$

porterebbe a risolvere il sistema

$$\frac{\delta\varphi}{\delta x_1} = -2\sum_{i=1}^{s} (x_{1j} - x_1)\; Y_i = 0\,, \quad \frac{\delta\varphi}{\delta x_2} = 0\,, \ldots, \frac{\delta\varphi}{\delta x_{s-1}} = 0\,,$$

la cui soluzione

$$x_1 = \frac{\sum\limits_{j=1}^{s} x_{1,j}\, Y_i}{\sum\limits_{j=1}^{s} Y_i}\,, \quad x_2 = \frac{\sum\limits_{j=1}^{s} x_{2,j}\, Y_j}{\sum\limits_{j=1}^{s} Y_i}\,, \ldots, \quad x_{s-1} = \frac{\sum\limits_{j=1}^{s} x_{s-1,j}\, Y_i}{\sum\limits_{j=1}^{s} Y_j}$$

è costituita dalle coordinate del baricentro dei punti $X_1\, X_2 \ldots X_s$ coi rispettivi pesi $Y_1\, Y_2 \ldots Y_s$. Si ritornerebbe, dunque, al concetto di media aritmetica secondo la sua interpretazione meccanica.

Quanto alla (2), scritta naturalmente nella forma

$$\psi(x_1\, x_2 \ldots x_{s-1}) = \sum_{j=1}^{s} |XX_j|\, Y_j = min.$$

essa esprimerebbe il problema di cercare nello spazio ad s-1 dimensioni in cui sono rappresentati $X_1\, X_2 \ldots X_s$ il punto di minima distanza da questi punti, quando ogni distanza sia ponderata col corrispondente valore di Y; problema che costituisce una generalizzazione del ben noto problema di FERMAT e di TORRICELLI, di determinare il punto avente la minima somma di distanze dai vertici di un triangolo.

Ma, ripetiamo, queste interpretazioni analitiche dei problemi espressi dalla (1) e dalla (2) non potrebbero ricevere conveniente applicazione nei riguardi delle serie sconnesse, e non è quindi il caso di soffermarsi più oltre su di esse.

Indici di mutabilità di una serie sconnessa

47. — *a*) Lo *scostamento semplice medio* ${}^{1}\eta_{A}$ *della serie data dalla* (o da una qualunque) *sua media aritmetica* $X_i = A$, sarà dato da

$$
{}^{1}\eta_{A} = \frac{\sum\limits_{j=1}^{i-1} | X_i X_j | \; Y_j + \sum\limits_{j=i+1}^{s} | X_i X_j | \; Y_j}{\sum\limits_{j=1}^{s} Y_i} = \frac{\sum\limits_{j=1}^{i-1} Y_j + \sum\limits_{j=i+1}^{s} Y_j}{\sum\limits_{j=1}^{s} Y_j} =
$$

$$
= \frac{\sum\limits_{j=1}^{s} Y_j - Y_i}{\sum\limits_{j=i}^{s} Y_j} = 1 - \frac{Y_i}{\sum\limits_{j=1}^{s} Y_j}
$$

dove deve essere ricordato che Y_i è il massimo (o uno dei massimi) dei valori di Y.

Si noti anche che, mentre per una serie ciclica di frequenze ${}^{1}\eta$ coincideva con ${}^{1}S_{R}$, per una serie sconnessa di frequenze l'indice ${}^{1}\eta_{A}$ non coincide con ${}^{1}S_{R}$ cioè con quello che già si dice (1) « scostamento semplice medio con ripetizione » della data serie.

b) Lo *scostamento quadratico medio* ${}^{2}\eta_{A}$ *dalla* (o da una qualunque) *sua media aritmetica* A, ha evidentemente l'espressione

$$
{}^{2}\eta_{A} = \left(\frac{\sum\limits_{j=1}^{i-1} (X_i X_j)^2 \, Y_j + \sum\limits_{j=i+1}^{s} (X_i X_j)^2 \, Y_i}{\sum\limits_{j=1}^{s} Y_i} \right)^{\frac{1}{2}}
$$

e poichè in questo caso delle serie sconnesse è sempre $(X_i X_j)^2 = 1 = | X_i X_j |$ (per $i \neq j$), così

$$
{}^{2}\eta_{A} = \left(\frac{\sum\limits_{j=1}^{s} Y_j - Y_i}{\sum\limits_{j=1}^{s} Y_j} \right)^{\frac{1}{2}} = \sqrt{{}^{1}\eta_{A}} \, .
$$

Anche qui si osservi che per una serie sconnessa di frequenze ${}^{2}\eta_{A}$ non coincide con ${}^{2}S_{R}$ cioè con quello che già si dice (2) « scostamento quadratico medio con ripetizione » della data serie.

(1) C. Gini. *Di una estensione del concetto di scostamento medio*, ecc. già citato.
(2) C. Gini. Loc. cit.

c) Lo *scostamento semplice medio dalla mediana* ${}^1\eta_M$ coincide ovviamente con lo scostamento analogo dalla media, cioè

$$ {}^1\eta_M = {}^1\eta_A \, . $$

Analogamente dicasi per lo scostamento quadratico medio dalla mediana: ${}^2\eta_M = {}^2\eta_A$. E si potrebbero altresì calcolare gli scostamenti medî, semplice e quadratico, dalla norma, cioè ${}^1\eta_N$, ${}^2\eta_N$.

d) Altri indici di mutabilità di una serie sconnessa sono già stati definiti nella Memoria testè citata, e, comunque, non sarebbe qui il luogo di parlarne diffusamente, dato che essi non dipendono nè si possono far dipendere dalla media o dalla mediana della data serie: essi sono la differenza semplice media, senza e con ripetizione Δ e Δ_R e la differenza quadratica media, senza e con ripetizione, ${}^2\Delta$ e ${}^2\Delta_R$.

V. — ANALOGIE E DIFFERENZE FRA LE SERIE RETTILINEE, CICLICHE E SCONNESSE

Applicabilità del calcolo alle diverse serie

48. — Se ci facciamo a comparare succintamente le conclusioni alle quali siamo pervenuti nello studio delle serie rettilinee, delle cicliche e delle sconnesse, troveremo, fra le une e le altre, analogie e diversità sulle quali vale la pena di soffermarsi, in quanto possono servire a meglio lumeggiare la struttura e l'essenza di ciascuna delle tre categorie.

Intanto si offre spontanea la constatazione che le serie rettilinee sono quelle alle quali più perfettamente possono aderire i procedimenti di calcolo, in dipendenza della loro struttura, che consente di rappresentare le modalità del carattere qualitativo X, in funzione del quale sono formate, mediante punti di una retta, o, si potrebbe anche dire, di uno spazio semplicemente connesso ad una dimensione; le diverse modalità di X si succedono in un ordine naturale, che può essere percorso in due sensi, e in ciascuno di questi sensi esiste una prima ed un'ultima modalità.

Le serie cicliche consentono anch'esse una rappresentazione grafica, quando si assuma come indice di ciascuna modalità di X un punto di una circonferenza, o più in generale di uno spazio ad una dimensione, doppiamente connesso; le diverse modalità si seguono

ancora in un ordine naturale che può essere percorso in due sensi; ma in nessuno di questi si può dire – all'infuori di una convenzione arbitraria – quali siano la prima e l'ultima modalità. Il trapasso da una ad un'altra modalità si può fare attraverso due distinte successioni di modalità intermedie. L'adattabilità del calcolo alle serie cicliche è minore che alle serie rettilinee, perchè appunto di quelle serie si può dare una figurazione meno semplice che di queste ultime.

Finalmente, delle serie sconnesse si può avere, attraverso le immagini delle s modalità di X, che vengono denotate dai vertici di un s-gono regolare nello spazio ad s-1 dimensioni, una rappresentazione teoricamente molto semplice, seppure scarsamente percettibile ai sensi: comunque, tali serie hanno, per definizione, la più semplice struttura possibile espressa dalla equidiversità delle modalità di X: e ciò è causa che anche i procedimenti di calcolo ad esse applicabili siano quanto mai semplici ed elementari.

Definizioni delle medie aritmetiche

49. — Quanto al modo di definire le usuali modalità medie abbiamo via via, per le varie specie di serie, sperimentato, ma non sempre con esito positivo, diverse definizioni, seguendo il concetto che quando una definizione tradizionale perdeva significato, si potesse surrogarla con una delle proprietà caratteristiche conseguenti dalla definizione stessa, in quei casi in cui essa non perde significato (principio di conservazione delle leggi formali).

Così dunque, assumendo come definizione iniziale di media aritmetica quella consueta e cioè:

a) modalità (valore) di X ottenuta dividendo per la somma dei valori di Y la somma dei prodotti dei valori di X pei corrispondenti valori di Y,

la quale è priva di significato per tutte le serie date in funzione di un carattere X qualitativo, le si surrogò, a seconda dei casi, una delle definizioni seguenti, le quali esprimono proprietà caratteristiche della media, quando la *a*) abbia significato:

b) modalità di X rispetto alla quale sia nulla la somma degli scostamenti delle altre modalità, opportunamente pesati;

c) modalità di X rispetto alla quale sia minima la somma dei quadrati degli scostamenti delle altre modalità, opportunamente pesati.

Dato, poi, che per una seriazione la media aritmetica secondo *a*) si può interpretare meccanicamente come il baricentro o come il centro di sospensione del sistema dei punti rappresentativi delle varie modalità, opportunamente pesati, o – ciò che torna lo stesso – come il punto di applicazione della risultante di più forze parallele (pesi o valori di Y) applicate ai punti di una retta (immagini delle modalità di X), così si surrogò infine alla definizione *a*) anche la

d) modalità di X avente la stessa espressione analitica del baricentro dei vari punti modalità, opportunamente pesati, o di un loro centro di sospensione, o della direzione della risultante di un conveniente sistema di vettori, di cui ciascuno rappresenti una modalità di X e il corrispondente valore di Y.

Come definizione di mediana si prese:

e) modalità $\overline{X}$ di X tale che la somma dei valori di Y corrispondenti alle modalità che precedono $\overline{X}$ sia minore della somma di tutte le altre modalità di Y, mentre la somma dei valori di Y corrispondenti alle modalità che non seguono $\overline{X}$, sia non minore della somma di tutti i rimanenti valori di Y (definizione tradizionale); e in surrogazione di questa:

f) modalità di X rispetto a cui sia minima la somma degli scostamenti assoluti dalle altre modalità, pesati mediante i valori di Y corrispondenti a tali modalità, (proprietà caratteristica che si verifica quando la *e*) abbia significato).

Come definizione di moda o norma si pose:

g) modalità di X alla quale corrisponde il massimo valore di Y (definizione tradizionale).

Stabilite tali definizioni si trovò che:

Per la media aritmetica:

1°) Le applicazioni delle *a*), *b*), *c*), *d*), alle serie rettilinee, che si investigano con perfetta analogia alle seriazioni, non appena alle modalità di X si siano fatti corrispondere numeri convenienti, conducono ad uno stesso ed unico risultato; ma perchè questo esista è generalmente necessario ampliare il concetto di modalità di X, introducendo altre modalità, di conto, oltre quelle effettive di X. (È ciò che accade anche per le seriazioni, in cui la media aritmetica, dato p. es. che i valori di X siano interi, non esiste generalmente tra i numeri interi). Si intende, tuttavia, che la *c*) potrebbe anche avere un significato, indipendentemente da qualunque ampliamento

nella successione delle modalità di X; ma si è, di proposito omessa tale forma di applicazione della *c*) perchè non suggerita nè dalla consuetudine nè da nessuna particolare ragione di utilità.

2°) Per una serie ciclica, rispetto a cui la *a*) non ha significato, le *b*) e *c*) forniscono multiplicità di risultati, ma tali risultati sono corrispondentemente gli stessi, purchè nella *c*) si intenda di considerare tutti i minimi relativi e non il solo o i soli minimi assoluti. La *d*) conduce ad un unico risultato, che non ha nulla in comune con quelli della *b*) e *c*) se la media si interpreta come definita dalla direzione della risultante di un sistema di vettori; conduce a due risultati (modalità opposte nel ciclo) se la media si interpreta come centro di sospensione di un sistema di punti ponderati sulla circonferenza, disposta in un piano verticale. Anche per le serie cicliche le *b*), *d*) esigono generalmente che si estenda la successione delle modalità di X mediante l'introduzione di altre modalità di conto; la *c*), invece, si potrebbe anche interpretare indipendentemente da tale estensione (modalità effettiva di X per cui sia minimo il momento secondo).

3°) Per una serie sconnessa le *a*), *b*) e *d*) non hanno potuto ricevere significato, mentre la *c*) ha trovato facile applicazione, rimanendo, naturalmente, nell'ambito delle modalità di X, dato che, come si è visto, non sarebbe neanche possibile ampliare la data serie con l'introduzione di altre modalità di conto.

Per la mediana:

1°) Ad una serie rettilinea è applicabile la definizione *e*); lo sarebbe anche la *f*) e fornirebbe lo stesso o gli stessi risultati della *e*); nè la *e*), nè la *f*) richiedono l'ampliamento dell'insieme delle modalità di X; possono, tuttavia, in taluni casi, considerarsi come mediane di una serie rettilinea, oltre a delle modalità effettive, anche tutte le modalità di conto comprese fra due modalità effettive.

2°) Ad una serie ciclica non è applicabile la *e*) ma soltanto la *f*); non è necessaria l'introduzione di modalità fittizie di X fra quelle effettive; non è escluso che tutte le infinite modalità di un certo intervallo si possano riguardare come mediane della serie.

3°) Ad una serie sconnessa è, similmente, applicabile soltanto la *f*), limitatamente, come sempre per tali serie, alle sole modalità effettive di X.

Finalmente *per la moda* si può conservare la definizione tradizionale *g*) per ciascuna delle tre categorie di serie.

Riassumendo:

per le serie rettilinee hanno significato le (a, b, c, d); (e, f,); g;

per le serie cicliche hanno significato le (b, c), d; f; g;

per le serie sconnesse hanno significato le c, f; g.

Sono rappresentate in parentesi quelle definizioni che danno luogo agli stessi risultati.

Si constata che le c) f) g) hanno significato per tutte le tre specie di serie (oltre che per le seriazioni); la b) e d) soltanto per le serie rettilinee e cicliche (oltre che per le seriazioni); le a) ed e) per le sole rettilinee (e per le seriazioni).

Indici di mutabilità indipendenti e dipendenti dalle medie

50. — Giova anche istituire un paragone analogo fra i diversi indici di mutabilità che si possono definire per le varie specie di serie. Intanto essi si possono distinguere in due classi:

Indici di mutabilità indipendenti dalle medie: 1°) differenza media senza e con ripetizione, semplice Δ e Δ_R, e quadratica ${}^2\Delta$ e ${}^2\Delta_R$ - in cui si intende per differenza fra due modalità di X il valore assoluto del loro scostamento; 2°) media ζ degli scostamenti algebrici (fra le modalità di X).

Indici di mutabilità dipendenti dalle medie: 3°) scostamento medio dalla media A (definita mediante a), o b), o c), semplice ${}^1\eta_A$ e quadratico ${}^2\eta_A$; 4°) scostamento medio dal centro di gravità o dalla risultante G o da un centro di sospensione (definizione d) semplice ${}^1\eta_G$ e quadratico ${}^2\eta_G$; 5°) scostamento medio dalla mediana M (definita mediante e) od f), semplice ${}^1\eta_M$ e quadratico ${}^2\eta_M$; 6°) scostamento medio dalla norma N (definizione g) semplice ${}^1\eta_N$ e quadratico ${}^2\eta_N$.

Ciascuno dei precedenti indici può essere definito per ciascuna delle diverse specie di serie, con unicità o multiplicità di risultati a seconda dei casi; meno che gli scostamenti medî, semplice e quadratico ${}^1\eta_G$ e ${}^2\eta_G$, che non hanno significato per le serie sconnesse; e precisamente:

la differenza media semplice e quella quadratica, senza e con ripetizione Δ, Δ_R , ${}^2\Delta$, ${}^2\Delta_R$ hanno, ciascuna, uno ed un solo valore in tutti i casi;

anche la media ζ degli scostamenti algebrici ha valore unico in ciascun caso; essa è nulla per le serie rettilinee e cicliche, e coincide con la Δ per le serie sconnesse;

gli scostamenti medî dalla media aritmetica A, semplice ${}^1\eta_A$ e quadratico ${}^2\eta_A$, hanno, ciascuno, valore unico per le serie rettilinee; e diversi valori (uno in corrispondenza a ciascuna media aritmetica) per le serie cicliche e sconnesse. (Non si devono confondere, per una serie ciclica, gli scostamenti medî *da una media* A, con quelli che abbiamo detto gli scostamenti medî *dalle medie*, semplice ${}^1\eta$ e quadratico ${}^2\eta$, i quali coincidono, invece, con quelli che si dicono scostamenti medî con ripetizione, semplice ${}^1S_R$ e quadratico ${}^2S_R$ della data serie ciclica).

Gli scostamenti medî, semplice ${}^1\eta_G$ e quadratico ${}^2\eta_G$ si possono definire, con unica determinazione, rispetto al baricentro per le serie rettilinee e rispetto alla risultante per le serie cicliche; e si possono invece definire con duplicità di risultato rispetto ai due centri di sospensione delle serie cicliche: che gli scostamenti medî siano disuguali rispetto a due tali centri, che sono punti opposti del ciclo, risulta da quanto si è detto al numero 35, circa il comportamento della $\Theta\,(X)$ per quanto riguarda lo scostamento semplice, e da quanto si è detto al numero 34 circa il momento secondo $\mu_2\,(X)$, per quanto riguarda lo scostamento quadratico. Si noti poi che ${}^1\eta_G$ e ${}^2\eta_G$ coincidono con gli scostamenti medî semplice e quadratico della media aritmetica, cioè con ${}^1\eta_A$, ${}^2\eta_A$ per le serie rettilinee; ed hanno invece valori diversi da quelli di tali indici per le serie cicliche.

Gli scostamenti medî, semplice e quadratico, dalla mediana ${}^1\eta_M$ e ${}^2\eta_M$ hanno, ciascuno, valore unico per le serie rettilinee e per le sconnesse, e possono avere pluralità di valori per le serie cicliche se di tali serie si considerano le mediane in senso lato (Cfr. n. 30); per le sconnesse ${}^1\eta_M$ e ${}^2\eta_M$ coincidono, naturalmente, con gli scostamenti medî, semplice e quadratico, delle medie aritmetiche.

Infine gli scostamenti medî, semplice ${}^1\eta_N$ e quadratico ${}^2\eta_N$, dalla norma possono avere uno o più valori, secondo che la serie sia uni- o plurimodale (1), per ciascuna delle tre specie di serie; e

(1) Per una serie rettilinea o ciclica si può parlare di massimo assoluto di Y e anche di massimo di Y relativo alle modalità *prossime ad una data modalità*; per una serie sconnessa si può parlare soltanto di massimo assoluto di Y; perciò in una serie plurimodale, rettilinea o ciclica, alle diverse mode potranno corrispondere diversi valori di Y; in una serie plurimodale sconnessa, alle diverse mode dovranno corrispondere valori uguali di Y.

per le serie sconnesse coincidono con gli scostamenti medî, semplice e quadratico, dalla media e dalla mediana.

Riassumiamo in uno specchio tutti questi risultati, non senza insistere sulla già segnalata circostanza che i diversi scostamenti semplici medî, e quadratici medî che si considerano non sono altro che particolari valori della funzione $\frac{\Theta(X)}{\Sigma Y_i}$ e rispettivamente della $\left(\frac{\mu_2(X)}{\Sigma Y_i}\right)^{\frac{1}{2}}$.

INDICI DI MUTABILITÀ		SERIE per cui possono definirsi e numero dei valori possibili			RELAZIONI fra diversi indici o loro particolari valori
Indipendenti dalle medie	Δ	*Rett.*, (1)	*Cicl.*, (1)	*Sconn.* (1)	
	${}^{2}\Delta$	R (1)	C (1)	S (1)	
	Δ_R	R (1)	C (1)	S (1)	
	${}^{2}\Delta_R$	R (1)	C (1)	S (1)	
	ζ	R (1)	C (1)	S (1)	Per R e per C: $\zeta = 0$ Per S: $\zeta = \Delta$
Dipendenti dalle medie	${}^{1}\eta_A$	R (1)	C (più)	S (più)	Per C: ${}^{1}\eta_A =\!|\!= {}^{1}\eta = {}^{1}S_R$
	${}^{2}\eta_A$	R (1)	C (più)	S (più)	Per C: ${}^{2}\eta_A =\!|\!= {}^{2}\eta = {}^{2}S_R$
	${}^{1}\eta_G$	R (1)	C (1)		Per R: ${}^{1}\eta_G = {}^{1}\eta_A$
	${}^{2}\eta_G$	R (1)	C (1)		Per R: ${}^{2}\eta_G = {}^{2}\eta_A$
	${}^{1}\eta_M$	R (1)	C (più)	S (più)	Per S: ${}^{1}\eta_M = {}^{1}\eta_A$
	${}^{2}\eta_M$	R (1)	C (più)	S (più)	Per S: ${}^{2}\eta_M = {}^{2}\eta_A$
	${}^{1}\eta_N$	R (più)	C (più)	S (più)	Per S: ${}^{1}\eta_N = {}^{1}\eta_M = {}^{1}\eta_A$
	${}^{2}\eta_N$	R (più)	C (più)	S (più)	Per S: ${}^{2}\eta_N = {}^{2}\eta_M = {}^{2}\eta_A$

VI. — SERIE DIPENDENTI DA PIÙ CARATTERI QUANTITATIVI O QUALITATIVI

Generalità

51. — Può darsi che in un insieme di fenomeni si considerino simultaneamente più caratteri $X, Y, \ldots Z, T$ e che si tenga conto delle modalità assunte da uno di essi, p. es. T, in corrispondenza ad ogni sistema di modalità assunte dai rimanenti caratteri $X, Y, \ldots Z$; si ha allora una « corrispondenza » nella quale $X, Y \ldots Z$ hanno un comportamento analogo a quello di variabili indipendenti, mentre T è paragonabile ad una funzione di $X, Y, \ldots Z$; dato, peraltro, che $X, Y, \ldots Z$ e T possono genericamente e promiscuamente essere caratteri quantitativi e qualitativi, esprimeremo la relazione fra $X, Y, \ldots Z$ e T scrivendo

$$T = S\,((X, Y, \ldots Z))$$

dove il simbolo di dipendenza di T da $X, Y, \ldots Z$ ricorda quello usato nell'analisi, ma non deve essere con esso confuso.

Per non dare alle considerazioni che seguono un carattere di soverchia astrazione e per attenerci alle condizioni di fatto più frequenti o più interessanti nella pratica, supporremo anzitutto che i caratteri considerati nei dati fenomeni siano dei caratteri statistici, che quelli aventi ruolo di variabili indipendenti siano soltanto due X ed Y, che T sia essenzialmente un carattere quantitativo e suscettibile, salvo avviso contrario, di soli valori non negativi, ed infine che le terne di corrispondenti modalità di X, Y, T effettivamente date siano in numero finito. Ciò non toglie che, occorrendo, il sistema delle coppie di modalità date per X ed Y possa, talora almeno, opportunamente ampliarsi con l'introduzione di altre coppie di modalità fittizie per X ed Y. Una corrispondenza così fatta si potrà dire una serie statistica a due dimensioni.

Quanto ai caratteri X ed Y essi potranno essere entrambi quantitativi e allora si avrà, più propriamente, a che fare con una *seriazione a due variabili;* o entrambi qualitativi e si avrà una *serie a due mutabili;* o infine uno X quantitativo ed uno Y qualitativo, e si avrà una *serie ad una variabile e a una mutabile.* Ma si intende che queste due classi di serie si potranno a loro volta distinguere in

sottoclassi, dipendentemente dalle diverse specie di mutabili che vi interverranno (rettilinee, cicliche, sconnesse).

Se la data seriazione o serie comprende s distinte coppie di modalità corrispondenti per X ed Y si potrà indicare genericamente una di tali coppie con $X_i\ Y_i\ (i = 1, \ldots s)$; ma ciò non significa che tutte le modalità di X e cioè $X_1\ X_2 \ldots X_s$ siano necessariamente distinte fra di loro, e similmente che lo siano le modalità $Y_1\ Y_2 \ldots Y_s$ di Y: in generale accadrà anzi che X ed Y siano suscettibili soltanto di t e rispettivamente di r modalità $(t \leq s,\ r \leq s)$. L'ordine di enumerazione delle coppie $X_i\ Y_i$ potrà essere arbitrario.

Se T_i denota il numero delle volte che nella data seriazione o serie si presenta la coppia di caratteri $X_i\ Y_i$, si avrà una seriazione o serie di frequenze assolute. In tal caso $\sum_{i=1}^{s} T_i = n$, numero totale dei casi considerati nella serie. Nelle stesse ipotesi, associando a ciascuna coppia $X_i\ Y_i$ la quantità $\frac{T_i}{\Sigma T_i}$ si avrà una seriazione o una serie di frequenze relative.

Seriazioni a due variabili

52. — Pensiamo anzitutto ad una immagine geometrica di tali seriazioni. È ovvio che, in riferimento ad un sistema cartesiano ortogonale di assi x, y, ogni coppia $(X_i\ Y_i)$ sarà rappresentata dal punto del piano di coordinate x_i, y_i, essendo x_i ed y_i le misure di X_i ed Y_i rispetto ad unità prescelte. Benchè ciò non sia essenziale, gioverà praticamente assumere unità tali che gli scostamenti massimi a cui danno luogo l'uno e l'altro carattere siano dello stesso ordine di grandezza e rappresentare poi le unità stesse con segmenti uguali sull'asse x e sull'asse y.

In conformità a questa rappresentazione è naturale che di due coppie di modalità $(X_i\ Y_i)$ ed $(X_j\ Y_j)$ si potrà dire scostamento, *e tale scostamento potrà considerarsi soltanto in valore assoluto*, la quantità $\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$.

Ciò posto, vediamo se e come possano stabilirsi i concetti di media aritmetica, di moda e di mediana.

Per la media aritmetica:

a) Si dice media aritmetica la coppia di modalità di X e Y rappresentata dal punto $(\bar{x}, \bar{y})$ del piano che ha come coordinate

le medie aritmetiche ponderate delle coordinate dei punti immagini delle coppie $(X_i\ Y_i)$, quando i pesi siano costituiti dai corrispondenti valori T_i; e cioè:

$$(1) \qquad \bar{x} = \frac{\sum_{i=1}^{s} x_i T_i}{\sum_{i=1}^{s} T_i}\,, \quad \bar{y} = \frac{\sum_{i=1}^{s} y_i T_i}{\sum_{i=1}^{s} T_i}\,;$$

il che significa, in sostanza, considerare separatamente le medie aritmetiche delle due seriazioni ad una dimensione:

$$\begin{cases} X_1\, X_2 \ldots X_s \\ T_1\, T_2 \ldots T_s \end{cases} \quad \text{ed} \quad \begin{cases} Y_1\, Y_2 \ldots Y_s \\ T_1\, T_2 \ldots T_s \end{cases}$$

La coppia $(\bar{x}, \bar{y})$ è invariante rispetto a qualunque trasformazione delle coordinate.

Una tale definizione non ha, generalmente, significato nello ambito delle coppie di modalità $(X_i\ Y_i)$ $(i = 1, \ldots s)$ effettivamente date; ha, invece, sempre significato e fornisce unicità di risultato considerando non i soli punti $(x_i\ y_i)$, ma anche gli altri punti del piano (rappresentativi di coppie di modalità di conto).

Del resto tale ampliamento è del tutto naturale, dato che i caratteri X ed Y sono quantitativi.

È poi importante osservare che la posta definizione non perde significato e conserva unicità di determinazione anche nel caso in cui non tutti i T_i siano positivi. Naturalmente in questa eventualità non si può più affermare (ciò che si vede facilmente sussistere se i T_i sono tutti non negativi) che il punto $(\bar{x}, \bar{y})$ denotante la media della serie sia interno al poligono convesso minimo contenente tutti gli $(x_i\ y_i)$ e avente come vertici alcuni di tali punti.

b) Se si dice media aritmetica la coppia di modalità di X ed Y rappresentata dal punto del piano rispetto a cui sia nulla la somma ponderata degli scostamenti dei punti che rappresentano le date coppie di modalità $(x_i\ y_i)$ (momento primo) cioè la somma $\mu_1(x, y) = \Sigma \sqrt{(x_i - x)^2 + (y_i - y)^2}\ T_i$, si vede subito che tale definizione non può avere significato, almeno nell'ipotesi posta che i valori T_i siano tutti non negativi, perchè gli scostamenti $\sqrt{(x_i - x)^2 + (y_i - y)^2}$ sono tutti positivi.

c) Si dice media aritmetica la coppia di modalità di X ed Y rappresentata dal punto $(\bar{y}, \bar{x})$ del piano rispetto a cui è minima la

somma ponderata dei quadrati degli scostamenti dei punti che rappresentano le date coppie di modalità (momento secondo), cioè

$$(2) \qquad \mu_2(x, y) = \sum_{i=1}^{s} \left\{ (x_i - x)^2 + (y_i - y)^2 \right\} T_i = min.$$

È evidente che tale definizione *può avere significato* (e fornisce uno o più risultati) *anche rimanendo nell'ambito delle sole coppie di modalità date* (X_i Y_i) (i = 1,...s) (significato in senso stretto).

Ma essa può altresì avere significato considerando, oltre quelli rappresentativi delle coppie (X_i Y_i), gli altri punti del piano, come si disse ad *a*). In tal caso la (2) fornisce, come si sa,

$$(2') \quad \frac{\partial(\mu_2)}{\partial x} = -2 \sum_{i=1}^{s} (x_i - x)\, T_i = 0 \quad , \quad \frac{\partial(\mu_2)}{\partial y} = -2 \sum_{i=1}^{s} (y_i - y)\, T_i = 0 \quad ,$$

e si ritorna alla definizione *a*) e quindi anche all'unicità del risultato. La prima delle (2′) significa che il momento primo della seriazione è nullo rispetto alla parallela ad y per $(\bar{x}, \bar{y})$; analogo significato ha la seconda; e poichè gli assi possono avere una direzione qualunque, si può dire che la media, secondo le definizioni *a*) e *c*) è rappresentata da un punto tale che il momento primo della seriazione data è nullo rispetto a qualunque retta, nel piano di rappresentazione, passante per il punto stesso.

Questa proprietà sostituisce, in certo modo, la *b*) (1).

Anche qui, va notato che la definizione continua ad essere applicabile se alcuni dei T_i sono negativi.

d) Se si assume come definizione di media aritmetica quella di baricentro dei punti (x_i y_i) coi pesi T_i, si ricade immediatamente nelle definizioni *a*) e *c*), per le quali si esige, dunque, la considerazione di punti del piano, oltre quelli (x_i y_i).

Concludendo, le definizioni *a*), *c*), *d*) danno sempre uno e lo stesso risultato, quando si amplî il campo delle modalità di X ed Y con le modalità di conto rappresentate dai punti del piano diversi dagli (x_i y_i); la definizione *c*) fornisce invece un risultato generalmente diverso dal precedente, quando venga applicata in senso stretto, cioè alle sole modalità effettive.

(1) Tale proprietà si può anche interpretare meccanicamente nel senso che la somma dei vettori T_i ($P_i - A$) [$P_i \equiv (x_i\, y_i)$, $A \equiv (\bar{x}, \bar{y})$] è nulla.

Per la mediana:

e) La definizione consueta, implicante l'esistenza di una prima e di un'ultima modalità del carattere in funzione del quale è costruita una serie, non può avere significato perchè le coppie di modalità $(X_i\ Y_i)$ in base a cui è data la nostra seriazione, non si seguono in un ordine determinato.

f) Se, invece, si dice mediana la coppia di modalità di X ed Y rappresentata dal punto $(\bar{x}, \bar{y})$ per cui è

$$(3) \qquad \Theta(x, y) = \sum_{i=1}^{s} \sqrt{(x_i - x)^2 + (y_i - y)^2}\, T_i = min.$$

intendendo di considerare nel radicale il solo valore positivo, tale definizione ha certamente significato e può avere pluralità di determinazioni anche nel solo insieme delle date coppie di modalità $(X_i\ Y_i)$.

La stessa definizione ha pure significato nell'insieme più ampio ottenuto con il considerare, oltre i punti corrispondenti alle coppie $(X_i\ Y_i)$, anche gli altri punti del piano (corrispondenti a coppie di modalità di conto). In questo caso la (3) esprime una generalizzazione del problema di Fermat e Torricelli, già ricordato al numero 44. La risoluzione della (3) conduce al sistema

$$(3') \qquad \begin{cases} \dfrac{\partial \Theta}{\partial x} = -\sum\limits_{i=1}^{s} \dfrac{(x_i - x)\, T_i}{\sqrt{(x_i - x)^2 + (y_i - y)^2}} = -\sum\limits_{i=1}^{s} T_i \cos \alpha_i = 0 \\[2ex] \dfrac{\partial \Theta}{\partial y} = -\sum\limits_{i=1}^{s} \dfrac{(y_i - y)\, T_i}{\sqrt{(x_i - x)^2 + (y_i - y)^2}} = -\sum\limits_{i=1}^{s} T_i\, sen\, \alpha_i = 0 \end{cases}$$

in cui le $\dfrac{\partial \Theta}{\partial x}$ e $\dfrac{\partial \Theta}{\partial y}$ sono indeterminate nei punti $(x_i\ y_i)$, e in cui α_i è l'angolo che la direzione da (x, y) ad $(x_i\ y_i)$ forma con la direzione positiva dell'asse x.

Perciò minimi della (3) cadranno generalmente in punti del piano distinti dagli $(x_i\ y_i)$ $(i = 1, \ldots s)$, (minimi di 1ª specie), ma non è escluso che ne cadano anche in punti del sistema $(x_i\ , y_i)$ (minimi di 2ª specie). Qui ci limitiamo ad osservare che pei minimi di prima specie le uguaglianze (3′) si debbono verificare comunque sia scelto l'asse X, e perciò, in senso meccanico, la somma degli s vettori, di moduli T_i e di argomenti α_i deve essere zero. Ciò porta

che i vettori stessi siano congruenti ai lati di un poligono chiuso e quindi, che per qualunque i, sia

$$T < T_1 + \ldots + T_{i-1} + T_{i+1} + \ldots + T_n \; (i = 1, 2, \ldots n) \; ;$$

se queste condizioni non sono soddisfatte, non potranno esistere se non minimi di 2ª specie.

Si osservi che, a differenza di quanto accadeva per la media aritmetica, la definizione di mediana di una seriazione a due dimensioni non può farsi dipendere da quella di mediana di una seriazione ordinaria. Difatti se delle seriazioni

$$\begin{cases} X_1 \, X_2 \ldots X_s \\ T_1 \;\, T_2 \ldots T_s \end{cases} \text{ed} \quad \begin{cases} Y_1 \, Y_2 \ldots Y_s \\ T_1 \;\, T_2 \ldots T_s \end{cases}$$

si considerassero le rispettive mediane $\overline{\overline{X}}$ ed $\overline{\overline{Y}}$, non si potrebbe affatto concludere che la coppia $(\overline{\overline{X}} \; \overline{\overline{Y}})$ costituisca la mediana della data seriazione a due dimensioni, nel senso dichiarato.

Per la norma:

Si può conservare ed ha validità la definizione consueta, e cioè quella di coppia di modalità $(X_i \; Y_i)$ a cui corrisponde il massimo od uno dei massimi valori di T; e ciò tanto se T_i sia la frequenza di $(X_i \; Y_i)$, come anche se T sia un altro carattere quantitativo.

Indici di variabilità:

Senza dilungarci ulteriormente nell'esame di seriazioni dipendenti da due caratteri quantitativi notiamo di sfuggita che per esse si potranno definire degli indici di variabilità, quasi tutti conformi a quelli che si possono calcolare per le seriazioni ordinarie. Usando, con analogia evidente di significato, i medesimi simboli già adoperati al numero 48, ci limiteremo ad osservare che:

Δ, Δ_R ${}^2\Delta$, ${}^2\Delta_R$ hanno, ciascuno, sempre un valore ed uno solo;

$\zeta = \Delta$, perchè, fra due coppie di modalità si può soltanto considerare lo scostamento assoluto;

${}^1\eta_A = {}^1\eta_G$, con unicità di valore;

${}^2\eta_A = {}^2\eta_G$, con unicità di valore;

${}^1\eta_M$ e ${}^2\eta_M$ possono avere più valori; e che, infine, anche ${}^1\eta_N$ e ${}^2\eta_N$ possono avere multiplicità di determinazioni.

Osservazione. — Per le seriazioni statistiche a due dimensioni, dipendenti cioè da due caratteri quantitativi e particolarmente per quelle costituite dalla popolazione di uno Stato, supposta distribuita

sopra un piano, di cui ogni punto risulti determinato dalle sue coordinate geografiche longitudine X e latitudine Y, è già in uso la considerazione di certe coppie di modalità che si dicono « centro di gravità » e « centro numerico o mediano » della seriazione e in particolare della popolazione considerata. Il « centro di gravità » è la coppia di modalità corrispondente alle nostre definizioni *a*), *c*), *d*). Invece la coppia detta « centro mediano » non corrisponde a nessuna delle definizioni sopra poste. Difatti questo centro, così come viene talora definito riferendosi alla distribuzione della popolazione di uno stato, è il punto di intersezione del meridiano e del parallelo tali che ciascuno di essi divida quella distribuzione in due parti ugualmente numerose.

Evidentemente il centro mediano così definito non soltanto manca della proprietà caratteristica della mediana che è quella di rendere minima la somma Θ degli scostamenti assoluti, ma non è neppure invariante rispetto al sistema di coordinate a cui si riferiscono i punti del piano: e cioè se questi punti si riferissero a coordinate diverse dalla latitudine, e dalla longitudine il cosidetto « centro mediano » verrebbe a mutare. La nostra definizione di mediana ha, invece, un significato obbiettivo per la distribuzione considerata.

Nel Vol. I, pag. 32 del *Fourteenth Census of the United States*, 1920, ripetendo un errore che già appare nel Vol. I, relativo al Censimento 1910 del medesimo Stato, è affermato: « If all the people « in the United States were to be assembled at one place, the center « of population (centro di gravità) would be the point wich they « could reach with the minimum aggregate travel, assuming that « they all travelled in direct lines from their residence to the meeting « place ». Non è affatto vero che il centro di gravità sia quello che gode di tale proprietà, perchè rispetto ad esso viene minimizzata, non già la somma degli scostamenti, ma la somma dei loro quadrati.

Il punto in cui la popolazione potrebbe riunirsi compiendo complessivamente il minimo tragitto possibile sarebbe precisamente quello che costituisce la mediana nel senso della definizione da noi assunta. Nel Rapporto citato (*ibid.*) si afferma successivamente: « This « would not be true of the median point », affermazione che è certamente giusta nei riguardi del « median point » così come è nel Rapporto stesso definito, ma che non lo è più rispetto alla mediana, così come deve essere giustamente intesa.

Serie ad una variabile e ad una mutabile rettilinea, o a due mutabili rettilinee

53. — Da quanto si disse a proposito delle serie rettilinee, relativamente alle quali si vide che esse sono assoggettabili a calcolo nello stesso modo delle seriazioni, per il fatto che esse dipendono da un carattere qualitativo le cui modalità si possono rappresentare mediante punti di una retta, si dedurrà subito che le serie in oggetto sono in tutto analoghe alle seriazioni dipendenti da due caratteri quantitativi, e che perciò si potrebbero per tali serie ripetere tutte le considerazioni fatte al numero precedente.

In sostanza, fatti corrispondere opportunamente dei numeri alle modalità della mutabile o delle due mutabili, si ricadrebbe dalla data serie ad una seriazione dipendente da due variabili; alle diverse modalità medie di tale seriazione corrisponderebbero modalità medie omonime per la serie; e gli indici di variabilità della seriazione sarebbero indici di mutabilità omonimi della serie.

Serie dipendenti da una variabile (o da una mutabile rettilinea) e da una mutabile ciclica

54. — Supponiamo, come sempre, che la mutabile ciclica X sia equispaziata e, nel solito modo, rappresentiamone le r possibili modalità mediante r punti di una circonferenza, a uguale distanza ciascuno dal successivo e l'ultimo dal primo. In corrispondenza a ciascuno P di tali punti-modalità rappresentiamo i possibili valori della variabile Y (o i numeri che si fanno corrispondere alle modalità della mutabile rettilinea) mediante quei punti della retta (passante per P ed ortogonale al piano della circonferenza) che hanno per ascisse quei valori. Sarà conveniente, se non essenziale, assumere come origine su tale retta il punto P e come unità di misura l'arco rettificato che segna lo scostamento assoluto fra due punti-modalità consecutivi di X. Inoltre, dal punto di vista pratico, gioverà anche scegliere relativamente ai due caratteri, delle unità di scostamento tali che gli scostamenti massimi fra le modalità dell'uno e e fra quelle dell'altro siano dello stesso ordine di grandezza. Così le date coppie di modalità $(X_i\ Y_i)(i = 1, 2, \ldots s)$ avranno come imma-

gini s punti giacenti sul cilindro avente per sezione retta l'accennata circonferenza.

Diremo «scostamento» di due coppie di modalità $(X_i\ Y_i)$ ed $(X_j\ Y_j)$ la lunghezza della geodetica minima congiungente M_i ed M_j immagini delle date coppie di modalità. Perciò, se lo scostamento assoluto delle modalità X_i e X_j del carattere ciclico X, è di p_{ij} gradi, e se y_i, y_j sono i valori corrispondenti alle modalità Y_i ed Y_j di Y, lo scostamento delle coppie $(X_i\ Y_i)$ ed $(X_j\ Y_j)$ sarà

$$\sqrt{p_{i,j}^2 + (y_j - y_i)^2}\ .$$

Tale scostamento, che può essere considerato soltanto in valore assoluto, è invariante per qualunque scorrimento del cilindro su se stesso. Ciò posto, è ancora possibile fissare i diversi concetti di media, assumendo fra le definizioni che già conosciamo, quelle che conservano un significato.

Senza troppo insistere su ciò, possiamo facilmente persuaderci che:

Per la media aritmetica:

La definizione *a*) non ha significato perchè alle modalità di Y corrispondono dei numeri o ascisse, ma non alle modalità di X.

Altrettanto dicasi della *b*), che esigerebbe la nozione di scostamento positivo e scostamento negativo.

La definizione *c*): modalità (coppia di modalità) per cui sia minima la somma ponderata dei quadrati degli scostamenti ha significato e, in generale, pluralità di determinazioni, sia limitandosi a considerare le sole coppie $(X_i\ Y_i)$ di modalità date, sia introducendo altre coppie di modalità di conto di X ed Y, rappresentate da punti della superficie cilindrica distinti da quelli che sono immagini delle $(X_i\ Y_i)$.

Si noti poi che se $Y_i = \text{cost} = 0$, la serie data si riduce ad una serie ciclica; e la definizione ora data di media aritmetica viene a coincidere con quella già data per le serie cicliche col criterio di rendere minimo il momento secondo.

Non ha significato la definizione *d*).

Per la mediana:

Non ha significato la definizione *e*), ed ha significato la *f*), sia limitatamente alle date coppie di modalità, sia considerando, come sopra, altre coppie di modalità fittizie.

Per la norma:

Conserva significato la definizione consueta.

Quanto agli indici di mutabilità, si vede subito che hanno, come sempre, uno ed un solo valore Δ, Δ_R, ${}^2\Delta_R$, ${}^2\Delta_R$; che $\zeta = \Delta$; che ${}^1\eta_A$ e ${}^2\eta_A$ possono, ciascuno, avere più valori, in corrispondenza alle diverse possibili medie aritmetiche; che ${}^1\eta_G$ e ${}^2\eta_G$ non hanno significato; che ciascuno degli indici ${}^1\eta_M$, ${}^2\eta_M$, ${}^1\eta_N$, ${}^2\eta_N$ può avere uno o più valori.

Osservazione. — Per avere una immagine geometrica dello scostamento di due coppie di modalità $(X_i\ Y_i)$ ed $(X_j\ Y_j)$ si pensi di sviluppare, sopra un piano e per tre volte di seguito, la superficie cilindrica sopra considerata, di modo che nella striscia di mezzo le s coppie di modalità saranno rappresentate dagli s punti Q_i $(i = 1, 2, \ldots s)$, essendo Q_i l'immagine di $(X_i\ Y_i)$, e nelle due striscie adiacenti, che denoteremo con gli indici I e II e che sono congruenti alla prima, le stesse coppie di modalità saranno rappresentate da altri due sistemi di s punti $Q_{I,i}$, $Q_{II,i}$ $(i = 1 \ldots s)$. Lo scostamento fra le due coppie $(X_i\ Y_i)$ ed $(X_j\ Y_j)$, misurato sulla geodetica minima fra i punti immagini di tali coppie sul cilindro, è rappresentato dal minimo dei due segmenti $Q_i\ Q'_j$, $Q_i\ Q''_j$, essendo Q'_j e Q''_j quei due punti, fra i tre Q_j, Q_{Ij}, Q_{IIj} giacenti sulle due generatrici che comprendono il punto Q_i.

Serie dipendenti da due caratteri ciclici

55. — Per avere un'immagine geometrica di una serie di questo tipo la prima idea che si affaccia alla mente è quella di assumere la superficie di un toro, su cui sia tracciata una rete di meridiani e di paralleli, e di rappresentare le modalità (che supponiamo, come sempre, equispaziate) del carattere ciclico X mediante meridiani equidistanti e le modalità (pure equispaziate) del carattere ciclico Y mediante paralleli condotti per punti equidistanti di un meridiano; ogni coppia $(X_i\ Y_i)$ sarebbe così rappresentata da un punto della superficie del toro, e come scostamento fra due coppie di modalità $(X_i\ Y_i)$ ed $(X_j\ Y_j)$ si potrebbe assumere la minima geodetica fra i punti immagini di quelle coppie.

Senonchè appare subito, per tale forma di rappresentazione, un inconveniente; il quale consiste in ciò, che se X_i ed X_{i+1} sono due modalità di X differenti di un grado e similmente Y_i ed Y_{i+1}, lo sco-

stamento tra le due coppie $(X_i\ Y_i)$ e $(X_{i+1}\ Y_i)$ risulta diverso dallo scostamento fra le coppie $(X_i\ Y_{i+1})$ ed $(X_{i+1}\ Y_{i+1})$, il che ripugna all'idea che noi abbiamo circa la struttura della data serie.

È vero che tale difficoltà si potrebbe superare adottando, sulla superficie del toro, una metrica conveniente; ma possiamo, più semplicemente, inspirandoci all'osservazione in fine del numero precedente, concepire una forma di rappresentazione che risulta immune da questi inconvenienti.

Si sviluppi sopra un piano, per tre volte successive, una superficie cilindrica su cui, mediante generatrici equidistanti, siano rappresentate le modalità equispaziate del carattere ciclico X; e similmente si sviluppi sullo stesso piano, per tre volte successive e in direzione ortogonale alla prima, una superficie cilindrica, p. es. di raggio uguale alla precedente, sulla quale un sistema di generatrici equidistanti rappresenti le modalità equispaziate di Y. La superficie comune ai due sviluppi sarà costituita da un quadrato diviso in nove quadrati (ciascuno di lato uguale alla circonferenza C dei cilindri); e le reti formate dai due sistemi di generatrici in questi quadrati saranno fra di loro congruenti; cosicchè distinguendo i quadrati circostanti a quello centrale con gli indici $I, II, \ldots VIII$, le s coppie di modalità $(X_i\ Y_i)$ avranno come immagini (s vertici entro ciascun quadrato, ossia nove configurazioni di punti Q_i, $Q_{\mathrm{I},i}$, $Q_{\mathrm{II},i} \ldots$ $Q_{\mathrm{VIII},i}$ $(i = 1, 2 \ldots s)$ fra loro congruenti. Ora, per definire lo scostamento delle coppie di modalità $(X_i\ Y_i)$ e $(X_j\ Y_j)$ si consideri l'immagine Q_i di $(X_i\ Y_i)$ nel quadrato centrale e quelle quattro, fra le nove immagini di $(X_j\ Y_j)$, che sono situate ai vertici del quadrato di lato C circonscritto a Q_i. Se quelle quattro immagini di $(X_j\ Y_j)$ si denotano con Q'_j, Q''_j, Q'''_j, Q''''_j si potrà dire «scostamento» delle coppie $(X_i\ Y_i)$ ed $(X_j\ Y_j)$ il minimo dei segmenti

$$Q_i Q'_j\,,\ Q_i Q''_j\,,\ Q_i Q'''_j\,,\ Q_i Q''''_j\ .$$

Una volta definito ciò che si può, senza ambiguità, riguardare come scostamento di due coppie di modalità, è chiaro che non vi saranno, almeno teoricamente, difficoltà a definire, anche per le serie dipendenti da due caratteri ciclici, le solite medie e i soliti indici di mutabilità.

Si vede poi, che per la media aritmetica le definizioni *a*), *b*), *d*) non possono ricevere significato; la *c*) invece, per la quale viene riguardata come media aritmetica la coppia di modalità rispetto a cui sia

minima la somma ponderata (mediante i T_i) dei quadrati degli scostamenti delle altre coppie di modalità, ha significato, e in generale multiplicità di determinazioni, sia in senso stretto (cioè considerando le sole coppie $X_i\ Y_i$ date), sia in senso lato. Per la mediana, delle definizioni *e*) ed *f*), potrà avere significato soltanto la *f*). Per la norma, potrà conservarsi la solita definizione *g*).

Conformemente a ciò, si potranno valutare gli indici di mutabilità Δ, Δ_R, ${}^2\Delta$, ${}^2\Delta_R$, ζ e sarà $\zeta = \Delta$; si potranno anche calcolare, e in generale con multiplicità di determinazioni, ${}^1\eta_A$, ${}^2\eta_A$, ${}^1\eta_M$, ${}^2\eta_M$, ${}^1\eta_N$, ${}^2\eta_N$; non potranno ricevere significato ${}^1\eta_G$ ${}^2\eta_G$.

È anche qui da notarsi l'opportunità che le unità degli scostamenti siano così scelte che risultino dello stesso ordine di grandezza gli scostamenti massimi delle modalità dell'uno e dell'altro carattere.

Serie dipendenti da due caratteri di cui uno almeno sconnesso

56. — Si possono pensare serie dipendenti da un carattere quantitativo (o qualitativo rettilineo) e da uno sconnesso; da uno ciclico e da uno sconnesso; da due sconnessi. Per questi tipi di serie non si può escludere la possibilità di definire delle medie e degli indici di mutabilità, ricorrendo all'applicazione di principî analoghi a quelli che abbiamo applicato per gli altri tipi, ma lo scarso interesse che le serie stesse presentano nella pratica, ci esonera dall'inoltrarci in considerazioni che sarebbero per necessità abbastanza complesse.

Chiudiamo questo studio col segnalare come, sia per la definizione delle modalità medie, sia per il calcolo degli indici di mutabilità abbia importanza basilare la determinazione di ciò che si deve intendere, in ciascun tipo di serie, per «scostamento» di due modalità (o di due sistemi di modalità) di quel carattere (o di quei caratteri) in relazione ai quali ciascuna serie è stata costruita.

VII. — APPLICAZIONI

Facciamo, ora, applicazione dei metodi esposti ad una serie di esempi offerti da rilevazioni statistiche concrete; ciò che, in particolare ci darà occasione a indagare comparativamente la portata dei risultati a cui quei metodi conducono.

Distribuzione dei matrimoni a seconda del giorno della settimana

57. — *a*) Come prima applicazione delle cose dette relativamente alle serie cicliche, consideriamo i numeri proporzionali a 10.000 dei numeri di matrimoni distribuiti per giorni della settimana, che si sono celebrati nei luoghi e tempi risultanti dal seguente prospetto (1); e delle serie cicliche così costituite determiniamo le medie aritmetiche nel senso delle definizioni *b*) e *c*), ($\Sigma\varepsilon = 0$ e $\Sigma\varepsilon^2 = min$).

(1) V. C. Gini. *Di una estensione del concetto di scostamento medio*, ecc., già citato.

NUMERI PROPORZIONALI A 10.000 DEI MATRIMONI CELEBRATI NEI DIVERSI GIORNI DELLA SETTIMANA

GIORNO DELLA SETTIMANA	MADRID	CREMONA	BERLINO	ROMA					
	1900-904	1887-89	1904	1877-86	1887-96	1897-906	1907-914	1914	1-6-1915 1-6-1917
Lunedì	1.301	2.742	1.146	609	1.119	1.641	1.740	1.762	1.512
Martedì	237	1.074	1.639	47	43	40	72	88	116
Mercoledì	1.235	455	1.190	943	1.191	1.003	1.005	1.079	1.226
Giovedì	1.487	1.694	1.631	2.593	2.360	2.210	2.079	2.122	2.214
Venerdì	583	96	1.027	30	16	30	53	88	75
Sabato	3.064	1.391	3.361	2.549	2.248	2.005	1.862	1.874	1.856
Domenica	2.093	2.548	6	3.229	3.023	3.071	3.189	2.987	3.001
TOTALI . . .	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000

In tutti questi casi la somma delle frequenze è $\Sigma\, Y_i = 10.000$ e quindi $\frac{s}{\Sigma Y_i} = 0{,}0007$. Poichè il numero delle modalità è dispari i punti opposti ai punti modalità bisecano gli archi minori determinati dalle coppie di modalità successive, e determinano a loro volta 7 intervalli. Riferendoci, p. es., alla serie dei matrimoni celebrati in Roma nel periodo 1877-86, la media aritmetica relativa all'intervallo $G'\,V'$, in cui si potrà assumere come origine L, avrà come scostamento da tale origine:

$$x_{\mathrm{I}} = \frac{\Sigma\, x_i\, Y_i}{\Sigma\, Y_i} = 0{,}0001 \times$$

$$\times\ (47.1 + 943.2 + 2593.3 - 30.3 - 2549.2 - 3329.1) = \mathbf{0{,}1295}\ .$$

Tale scostamento determina una media ordinaria, cioè compresa nell'intervallo stesso $G'\,V'$.

Procedendo nel senso positivo, gli scostamenti delle altre medie dalla stessa origine L saranno, conformemente alla seconda regola data:

Per attraversare	V':	x_{II}	$= 0{,}1295 + 30 \times 0{,}0007 = 0{,}1505$
» »	S':	x_{III}	$= 0{,}1505 + 2549 \times 0{,}0007 = \mathbf{1{,}9348}$
» »	D':	x_{IV}	$= 1{,}9348 + 3229 \times 0{,}0007 = 4{,}1951$
» »	L':	x_{V}	$= 4{,}1951 + 609 \times 0{,}0007 = 4{,}6214$
» »	Ma':	x_{VI}	$= 4{,}6214 + 47 \times 0{,}0007 = \mathbf{4{,}6543}$
» »	Me':	x_{VII}	$= 4{,}6543 + 943 \times 0{,}0007 = 5{,}3144$

Prova: Si attraversa G': $x_{\mathrm{I}} = 5{,}3144 + 2593 \times 0{,}0007 = 7{,}1295$.

Per vedere quali fra queste medie siano contenute nei rispettivi intervalli, riduciamo lo scostamento di x_{II} all'origine Ma (punto di mezzo dell'intervallo corrispondente $V'\,S'$), quello di x_{III} alla origine Me (punto di mezzo di $S'\,D'$), ecc. Gli scostamenti dalle nuove origini saranno:

x_{I}	x_{II}	x_{III}	x_{IV}	x_{V}	x_{VI}	x_{VII}
0,1295	—0,8495	**—0,0652**	1,1951	—0,6214	**—0,3457**	0,6856

Perciò, le tre sole medie aritmetiche espresse dagli scostamenti $x_{\mathrm{I}} = 0{,}1295$, $x_{\mathrm{III}} = -0{,}0652$, $x_{\mathrm{VI}} = -0{,}3457$ (che non superano in valore assoluto mezza unità) sono contenute nei rispettivi intervalli $G'\,V'$, $S'\,D'$, $Ma'\,Me'$, mentre gli altri scostamenti rappresentano medie fittizie, cioè punti esterni agli intervalli a cui competono.

Non esiste nessuna media limite e nessuna media *sui generis*. Le sole medie accettabili sono dunque quelle rappresentate da x_{I}, x_{III}, x_{VI} (medie ordinarie).

In modo analogo sono stati calcolati, sempre rispetto alla origine L e si sono raccolti nella seguente tabella, gli scostamenti delle medie accettabili delle altre serie cicliche date, nessuna delle quali ammette medie limite o medie *sui generis*.

Medie aritmetiche ordinarie misurate dal lunedì

	INTERVALLI	Madrid	Cremona	Berlino	Roma					
		1900-904	1887-89	1904	1877-86	1887-96	1897-906	1907-914	1914	1-6-1915 1-6-1917
x_{I}	$G' \ldots V'$	— 0.2802	0.1448	— 0.0897	0.1295	0.1938	0.1505	0.1247	0.1613	0.2272
x_{II}	$V' \ldots S'$	..	..	0.6292	..	..	..	..	..	..
x_{III}	$S' \ldots D'$	2.2727	..	..	1.9348	1.7786	1.5750	..	1.5347	1.5789
x_{IV}	$D' \ldots L'$	..	2.9693	2.9861	..	..	..	..	..	..
x_{V}	$L' \ldots Ma'$	..	..	3.7883	..	..	..	..	..	..
x_{VI}	$Ma' \ldots Me'$	4.8144	..	4.9356	4.6543	4.7081	4.9014	4.9659	4.9206	4.8192
x_{VII}	$Me' \ldots G'$	5.6789	5.9590	5.7686	..	5.5418	5.6035	5.6694	5.6759	5.6774

Fra le serie cicliche considerate, quella relativa a Berlino 1904 ammette il maggior numero di medie aritmetiche ordinarie, ciò che significa una maggiore simmetria nella distribuzione dei matrimoni lungo il corso della settimana. Per tutte le serie l'intervallo *G' V'* (di cui *L* è il punto di mezzo) ammette una media aritmetica ordinaria. La sola serie di Berlino 1904 ammette una media aritmetica ordinaria nell'intervallo *V' S'*.

Riferendoci ancora alla serie relativa a Roma 1877-86, e considerandone tutte le medie aritmetiche, accettabili e non, vediamo che gli scostamenti di tali medie dalle modalità che cadono negli intervalli a cui le medie stesse corrispondentemente competono sono:

da	L	Ma	Me	G	V	S	D
a	x_{I}	x_{II}	x_{III}	x_{IV}	x_{V}	x_{VI}	x_{VII}
	0,1295	0,1505−1	1,9348−2	4,1951−3	4,6214−4	4,6543−5	5,3144−6

Perciò (n. 39) lo scostamento semplice medio dalle medie ${}^1\eta$ (coincidente, come sappiamo, con quello che era già stato definito dal Gini come scostamento semplice medio con ripetizione della serie, ${}^1S_{\mathrm{R}}$) e lo scostamento quadratico medio dalle medie ${}^2\eta$ (coincidente con lo scostamento quadratico medio con ripetizione della serie, ${}^2S_{\mathrm{R}}$) sono rispettivamente:

$$ {}^1\eta = {}^1S_{\mathrm{R}} = \frac{1}{10000}\begin{pmatrix} 609 \, . \, 0{,}1295 + 47 \, . \, |0{,}1505 - 1| + 943\,|1{,}9348 - \\ - 2| + 2593\,|4{,}1951 - 3| + 30\,|4{,}6214 - 4| + \\ + 2549\,|4{,}6543 - 5| + 3229\,|5{,}3144 - 6| \end{pmatrix} = $$
$$ = 0{,}6393 $$

$$ {}^2\eta = {}^2S_{\mathrm{R}} = \left\{ \frac{1}{10000} \begin{pmatrix} 609 \, . \, 0{,}1295^2 + 47\,(0{,}1505 - 1)^2 + \\ + \, . \; . \; . \; . + 3229\,(5{,}3144 - 6)^2 \end{pmatrix} \right\}^{\frac{1}{2}} = 0{,}7474 $$

(Per constatare la coincidenza di questi coi valori già calcolati dal Gini, cfr. *Di una estensione ecc.*, già cit.). Non ci soffermiamo a ricalcolare i valori analoghi per le altre serie.

b) Vogliamo, ora, delle medesime serie cicliche, determinare le medie aritmetiche nel significato meccanico. Considerate le date frequenze dei matrimoni come moduli di vettori uscenti dal centro e passanti per quei punti di una circonferenza che rappresentano le modalità del carattere ciclico «giorno della settimana», la media aritmetica nel significato meccanico sarà rappresentata dall'anomalia della risultante di tali vettori. Riferendosi a un sistema di assi ortogonali ξ, η condotti pel centro della stessa circonferenza e dei

quali ξ passi per il punto modalità L, si avrà che le componenti dei detti vettori su tali assi sono, ad es. per la serie relativa a Roma 1877-86:

MODALITÀ	ANOMALIA rispetto a ξ	FREQUENZE Roma 1877-86	COMPONENTI su ξ	COMPONENTI su η
L	0	609	609 —	0
Ma	$\frac{2\pi}{7}$	47	29,30	36,75
Me	$\frac{4\pi}{7}$	943	— 209,84	919,36
G	$\frac{6\pi}{7}$	2.593	— 2.336,20	1.125,06
V	$\frac{8\pi}{7}$	30	— 27,03	— 13,02
S	$\frac{10\pi}{7}$	2.549	— 567,21	— 2.485,09
D	$\frac{12\pi}{7}$	3.229	2.013,25	— 2.524,54
		10.000	— 488,73	— 2.941,48

Il modulo della risultante di tali vettori è dunque:

$$R = \sqrt{488,73^2 + 2941,48^2} = 2981,80$$

e quindi la sua anomalia rispetto a ξ è, tenuto conto dei segni delle componenti:

$$\varphi = arcos \frac{-488,73}{2981,80} = 260^{\circ}34'$$

Condotto analogamente il calcolo per tutte le altre serie cicliche considerate, si trovano i seguenti risultati:

	Madrid	Cremona	Berlino	Roma					
				1877-86	1887-96	1897-906	1907-14	1914	1-6-1915 1-6-1917
R	2842,80	3277,57	1364,22	2981,80	2346,65	2561,50	2683,88	2392,60	2066,52
φ	268°38′	335°16′	205°4′	260°34′	273°3′	290°24′	296°54′	295°31′	290°1′

La risultante di modulo minimo è quella che si riferisce alla serie relativa a Berlino, ciò che conferma la maggiore simmetria con cui i matrimoni di questa serie si distribuiscono nel corso della settimana rispetto a quelli delle altre serie; maggiore simmetria di cui si era avuto indizio anche dal maggior numero di medie aritmetiche nel senso della prima definizione. Poichè ogni modalità viene rappresentata dal punto di mezzo di un arco avente l'ampiezza $2\pi/7$, così si vede che delle tre serie di Madrid, Cremona Berlino, le medie in senso meccanico sono rispettivamente prossime al S, alla D, al V. Delle sei serie che si riferiscono a Roma le medie in senso meccanico cadono per le prime due poco oltre il S, per le altre quattro poco prima della D, come se in progresso di tempo vi fosse stato uno spostamento complessivo dei matrimoni nel senso che va dal S alla D. La serie di Madrid, più di quelle di Cremona e Berlino, mostra una certa analogia, sia per il valore assoluto che per l'anomalia della risultante, con quelle di Roma. La serie di Cremona dà luogo al massimo valore assoluto della risultante, il che significa una maggiore asimmetria di tale serie rispetto alle altre, in accordo con l'esistenza di un numero piuttosto piccolo (3) di medie aritmetiche nel senso della prima definizione, e con la concentrazione di un grande numero di matrimoni in un solo giorno (domenica): tale concentrazione è la massima fra tutte quelle che si rilevano nelle serie date.

Come risulta da questo esempio, le conclusioni che si traggono dall'esame delle medie aritmetiche nel senso della prima definizione si accordano con quelle deducibili dall'esame delle risultanti e delle anomalie, ma non hanno esattamente la stessa portata. Le medie aritmetiche nel senso della prima definizione danno, ciascuna, indizio di una certa simmetria della serie rispetto ad essa, e quindi,

a parità di modalità effettive, un maggior numero di medie aritmetiche accettabili denota una maggiore simmetria della serie lungo il ciclo. Ciascuna media aritmetica, nel senso detto, sintetizza tutta la serie ciclica, ma la sintesi offerta da una di esse non può generalmente ritenersi equivalente a quella offerta da un'altra, tanto è vero che, come si è già osservato, gli indici di variabilità ${}^{1}S_{A}$, ${}^{2}S_{A}$ deducibili da una di esse non coincidono con quelli relativi ad una altra media. La posizione delle medie aritmetiche secondo la prima definizione è subordinata alla convenzione fatta che gli scostamenti fra le modalità del ciclo non possano, in valore assoluto, superare mezzo ciclo.

Anche la considerazione del valore assoluto e dell'anomalia della risultante può indicare la maggiore o minore simmetria complessiva delle serie considerate, poichè, sempre supponendo un ugual numero di modalità effettive, sarà maggiore la simmetria là dove è minore il valore assoluto della risultante. Quanto all'anomalia della risultante stessa, essa sintetizzerà la serie ciclica, indipendentemente dalla particolare convenzione ora ricordata. Inoltre la simultanea considerazione del valore assoluto e dell'anomalia delle risultanti di più serie cicliche potrà essere significativa anche se nelle serie stesse il carattere ciclico da cui dipendono sia stato eventualmente distinto in modalità diversamente numerose da serie a serie (Cfr. n. 5), mentre in questa eventualità non sarebbe legittima la comparazione fra i corrispondenti sistemi delle medie aritmetiche nel senso della prima definizione.

c) Infine, determiniamo le mediane delle medesime serie. Sappiamo (n. 30) che comunque sia la parità del numero delle modalità della mutabile da cui dipende la serie, le mediane cadono in corrispondenza a tali modalità e perciò prescindendo dalla determinazione della somma $\Theta(X)$ degli scostamenti assoluti in tutti i punti del ciclo, le mediane si possono calcolare valutando la $\Theta(X)$ soltanto in corrispondenza alle dette modalità e osservando dove cadano il minimo od i minimi della somma stessa $\Theta(X)$. Per Roma 1877-86 si trova:

in	L	Ma	Me	G	V	S	D
$\Theta(X) =$	18129	20933	21252	17679	15454	**12633**	13920

e si conclude, pertanto, che l'unica mediana è S. Gli scostamenti semplici e quadratici da tale mediana sono rispettivamente:

$${}^{1}\eta_{M} = 1{,}2633 \qquad {}^{2}\eta_{M} = 1{,}5804$$

Similmente si procede per le altre serie; e si trovano i risultati raccolti nel seguente prospetto:

	Madrid	Cremona	Berlino	Roma					
				1877-86	1887-96	1897-906	1907-14	1914	1-6-1915 1-6-1917
Mediana	*S*	*L*	*S*	*S*	*S*	*D*	*D*	*D*	*D*
${}^{1}\eta_{M}$	1,2668	1,2684	1,5074	1,2633	1,3699	1,3425	1,3104	1,3591	1,4070
${}^{2}\eta_{M}$	1,6454	1,6467	1,9391	1,5804	1,6751	1,8123	1,7849	1,8207	1,8733

In tutte queste serie, meno che per quella di Berlino, l'unica mediana di poco si discosta dalla modalità corrispondente all'anomalia della risultante cioè dalla media in senso meccanico. Poichè la mediana di una serie ciclica di frequenze ha la proprietà di rendere minima la somma degli scostamenti assoluti, così si intende che essa cadrà in quella zona del ciclo in cui si addensa il maggior numero dei casi. Altrettanto è da presumersi che generalmente accada per la media in senso meccanico. Perciò si capisce come mediana e media in senso meccanico debbano, generalmente, avere posizioni prossime sul ciclo. Nel caso di simmetria rispetto alla media in senso meccanico, questa costituisce anche la mediana delle serie. La divergenza fra queste due medie potrà quindi assumersi come un indice della asimmetria della serie rispetto alla media in senso meccanico, e in questo senso si potrà dire che la asimmetria della serie relativa a Berlino, rispetto alla sua media in senso meccanico, è maggiore di quella delle altre serie.

PRECIPITAZIONI MENSILI OSSERVATE IN ROMA

58. — *a*) In secondo luogo consideriamo la quantità d'acqua caduta in Roma in ciascun mese dei periodi sotto indicati (1).

QUANTITÀ DELLE PRECIPITAZIONI MENSILI OSSERVATE IN ROMA PER CIASCUN PERIODO SOTTO INDICATO (ESPRESSE IN mm.)

MESE	ANNI 1913-1915	ANNI 1918-1921	ANNI (1913-15) + + (1918-21)
Gennaio	337	210	547
Febbraio	433	158	581
Marzo	194	259	453
Aprile	242	336	578
Maggio	282	182	464
Giugno	176	194	370
Luglio	30	19	49
Agosto	109	104	213
Settembre	184	183	367
Ottobre	306	529	835
Novembre	332	287	619
Dicembre	433	304	737
	3.048	2.765	5.813

Determiniamo le medie aritmetiche nel senso delle definizioni basate sulla considerazione della somma degli scostamenti e su quella della somma dei quadrati degli scostamenti.

I punti opposti ai punti modalità coincidono con punti modalità e determinano successivamente 12 intervalli. Per la serie corrispondente alla prima colonna la media relativa all'intervallo *Lug'*...

(1) Dati ricavati dall'« Annuario Statistico Italiano ».

Ag', riferita al punto di mezzo O di questo intervallo, avrà come scostamento da tale origine:

$$x_{\mathrm{I}} = \frac{1}{3048}\left[\frac{1}{2}(423-337) + \frac{3}{2}(194-433) + \frac{5}{2}(242-332) + \right.$$
$$\left. + \frac{7}{2}(282-306) + \frac{9}{2}(176-184) + \frac{11}{2}(30-109)\right] = -\frac{365}{1016} =$$
$$= -\mathbf{0{,}359}$$

Tale scostamento determina una media ordinaria.

Procedendo nel senso positivo gli scostamenti delle altre medie dalla stessa origine O saranno, poichè $\frac{s}{\Sigma Y_{\mathrm{I}}} = \frac{12}{3048} = \frac{4}{1016}$:

Per attrav. Ag′: $x_{\mathrm{II}} = \frac{365}{1016} + \frac{4\times 109}{1016} = \frac{71}{1016} = 0{,}070$

» » Sett′: $x_{\mathrm{III}} = \frac{71}{1016} + \frac{4\times 184}{1016} = \frac{807}{1016} = 0{,}794$

» » Ott′: $x_{\mathrm{IV}} = \frac{807}{1016} + \frac{4\times 306}{1016} = \frac{2031}{1016} = 1{,}199$

» » Nov′: $x_{\mathrm{V}} = \frac{2031}{1016} + \frac{4\times 332}{1016} = \frac{3359}{1016} = 3{,}306$

» » Dic′: $x_{\mathrm{VI}} = \frac{3359}{1016} + \frac{4\times 433}{1016} = \frac{5091}{1016} = \mathbf{5{,}011}$

» » Genn′: $x_{\mathrm{VII}} = \frac{5091}{1016} + \frac{4\times 337}{1016} = \frac{6439}{1016} = \mathbf{6{,}338}$

» » Febb′: $x_{\mathrm{VIII}} = \frac{6439}{1016} + \frac{4\times 423}{1016} = \frac{8131}{1016} = 8{,}003$

» » Mar′: $x_{\mathrm{IX}} = \frac{8131}{1016} + \frac{4\times 194}{1016} = \frac{8907}{1016} = 8{,}767$

» » Apr′: $x_{\mathrm{X}} = \frac{8907}{1016} + \frac{4\times 242}{1016} = \frac{9875}{1016} = 9{,}719$

» » Magg′: $x_{\mathrm{XI}} = \frac{9875}{1016} + \frac{4\times 282}{1016} = \frac{11003}{1016} = 10{,}830$

» » Giug′: $x_{\mathrm{XII}} = \frac{11003}{1016} + \frac{4\times 176}{1016} = \frac{11707}{1016} = 11{,}523$

Prova: Per attrav. Lug′: $x_{\mathrm{I}} = \frac{11707}{1016} + \frac{4\times 30}{1016} = \frac{11827}{1016} = 12 - \frac{365}{1016}$

Sono accettabili le sole medie x_{I}, x_{VI}, x_{VII} le quali, poichè cadono entro i rispettivi intervalli, costituiscono delle medie ordinarie. Non esiste nessuna media *sui generis* e nessuna media limite. Similmente si conduce il calcolo per le altre due serie cicliche. Gli scostamenti da O di tutte le medie accettabili, che anche per le altre due serie sono esclusivamente medie ordinarie, sono raccolte nel seguente prospetto.

MEDIE ARITMETICHE ORDINARIE MISURATE DAL PUNTO O

INTERVALLI	1913-15	1918-21	(1913-15) + + (1918-21)
Lug'...Ag'	$x_{I} = -0{,}359$		$x_{I} = -0{,}464$
Ag'...Sett'.			
Sett'...Ott'			
Ott'...Nov'		$x_{IV} = 2{,}961$	
Nov'...Dic'		$x_{V} = 4{,}207$	$x_{V} = 3{,}735$
Dic'...Genn'	$x_{VI} = 5{,}011$		$x_{VI} = 5{,}256$
Genn'...Febb'	$x_{VII} = 6{,}338$	$x_{VII} = 6{,}438$	$x_{VII} = 6{,}385$
Febb'...Mar'.		$x_{VIII} = 7{,}124$	
Mar'...Apr'		$x_{IX} = 8{,}248$	
Apr'...Magg'.			
Magg'...Giug'		$x_{XI} = 10{,}496$	
Giug'...Lug'		$x_{XII} = 11{,}338$	$x_{XII} = 11{,}435$

Il lettore potrà facilmente rappresentare sopra una circonferenza, per ciascuna delle tre serie, le posizioni delle rispettive medie aritmetiche. La serie relativa al triennio 1913-15 è la più asimmetrica, poichè ammette soltanto tre medie accettabili su 12 possibili; e la più simmetrica è quella per il quadriennio 1918-21. L'intervallo *Genn'... Febb'*, coincidente con *Lug...Ag*, contiene una media ordinaria per ciascuna delle tre serie. Nei tre intervalli *Ag'...Sett'*, *Sett'...Ott'*, *Apr'...Magg'* non cadono medie aritmetiche per nessuna delle tre serie.

Relativamente alle medie aritmetiche della serie 1913-15 gli scostamenti medî, semplici e quadratici, sono rispettivamente:

da	$_1\eta$	$_2\eta$
x_{I} . . .	2,32	2,78
x_{VI} . . .	3,65	3,98
x_{VII} . . .	3,59	3,92

b) Determiniamo ora le risultanti delle tre serie prese in esame.

Per la prima il calcolo della risultante può farsi come appare dal seguente prospetto, in cui è da intendersi che le direzioni ξ e η

sulle quali vengono determinate le componenti di ciascun vettore, siano quelle che dal centro della circonferenza su cui si rappresentano i diversi mesi vanno, rispettivamente, ai punti rappresentativi del gennaio e dell'aprile. I mesi si suppongono tutti della stessa durata, di modo che la serie si considera come equispaziata, e le direzioni ξ e η sono ortogonali; si suppone inoltre che il senso positivo delle rotazioni sia quello definito dal minimo arco che va da gennaio a febbraio.

MODALITÀ	Anomalia rispetto a ξ	Precipitazioni mensili in mm. 1913-1915	Componenti su ξ	Componenti su η
Gennaio	0°	337	337	0
Febbraio	30°	423	366,33	211,5
Marzo.	60°	194	97	168,01
Aprile	90°	242	0	242
Maggio	120°	282	— 141	244,22
Giugno	150°	176	— 152,42	88
Luglio	180°	30	— 30	0
Agosto	210°	109	— 94,40	— 54,5
Settembre	240°	184	— 92	— 159,35
Ottobre.	270°	306	0	— 306
Novembre	300°	332	166	— 287,52
Dicembre	330°	433	374,99	— 216,5
		(1) 3048	(2) 831,50	(3) — 70,14

(1) Precipitazione totale.
(2) Componente della risultante su ξ.
(3) Componente della risultante su η.

$$\text{Modulo } R \text{ della risultante} = \sqrt{831{,}50^2 + 70{,}14^2} = 834{,}45$$

$$\text{Anomalia } \varphi \text{ della risultante} = \mathit{arctg}\,\frac{-70{,}14}{831{,}50} = -4^\circ 49' 10'' \,.$$

Se si eseguiscono i calcoli analoghi per le altre due serie e si raccolgono i risultati in una sola tabella si ottiene:

RISULTANTI DELLE PRECIPITAZIONI MENSILI OSSERVATE IN ROMA NEI PERIODI SOTTO INDICATI

PERIODO	R	φ	PRECIPITAZIONI TOTALI in mm.
1913-1915	mm. 834,45	— 4°49′10″	3.048
1918-1921	mm. 489,84	— 30°9′36″	2.765
(1913-15) + (1918-21) . . .	mm. 1255,03	— 14°9′18″	5.813

Le anomalie delle tre risultanti determinano sulla circonferenza rappresentativa del carattere X « mese dell'anno » tre punti che sono le immagini delle medie aritmetiche delle serie date, nel senso meccanico stabilito. Poichè ogni modalità di X viene figurata mediante un punto che è il centro di un arco di 30° corrispondente alla durata di un mese, così le anomalie di — 4°49′10″ e — 14°9′18″ rappresentano modalità del ciclo che cadono nell'ambito del mese di gennaio; mentre l'anomalia — 30°9′36″ corrisponde ad una modalità che cade nell'ambito del dicembre e verso la sua metà.

È quasi superfluo osservare che i moduli delle tre risultanti non sono in relazione con la sola piovosità complessiva di ciascuno dei periodi considerati, poichè dipendono anche dalle anomalie delle varie modalità del carattere X « mese dell'anno », cosicchè a periodi di uguali piovosità complessive possono ben corrispondere risultanti dotate di moduli disuguali.

Nelle tre serie dell'esempio considerato l'elemento che presenta una maggiore stabilità è la media aritmetica nel senso meccanico cioè l'anomalia della risultante, la quale anomalia è sempre negativa e denotante una modalità di X compresa fra i mesi consecutivi dicembre-gennaio.

La circostanza, poi, che nelle tre serie tale anomalia si mantiene negativa, dimostra una preponderanza del secondo semestre dell'anno

rispetto al primo nella formazione della detta risultante (si osservi la maggiore piovosità complessiva degli ultimi tre mesi dell'anno rispetto ai primi tre).

È pure da osservarsi, in riferimento alla serie 1913-15, che i valori minimi degli scostamenti medî, semplice e quadratico, si verificano rispetto a quella media aritmetica X_1 (nel senso della prima definizione) che è più prossima alla media in senso meccanico.

Regime dei venti nel villaggio Duca degli Abruzzi

59. — *a*) Il procedimento di ricerca della media aritmetica in senso meccanico (modalità corrispondente alla anomalia della risultante) appare particolarmente indicato per sintetizzare il regime dei venti spirati in un certo luogo e periodo di tempo. Applichiamolo, dunque, come esempio, alle osservazioni anemometriche compiute alle 6^h, alle 12^h e alle 18^h durante l'anno 1922, nell'Osservatorio meteorologico del Villaggio S. A. R. Duca degli Abruzzi (1) (Lat. 2°41′45″ *N*, Long. 45°29′34″ *E* Gr.), tenendo conto della velocità oraria del vento in chilometri (2).

Nella tabella che segue sono indicate per la serie delle osservazioni a 6^h, e separatamente per ciascuno dei mesi, le direzioni dei venti, le loro anomalie rispetto alla direzione fondamentale *W-E*, le somme delle velocità osservate in ciascun mese e in ciascuna direzione, le componenti di ciascuna di tali somme nelle direzioni *W-E* e *S-N*, ottenute moltiplicando quelle somme per il coseno e per il seno della rispettiva anomalia; e infine le somme di tutte le componenti nelle direzioni stesse.

(1) D. Omodei e F. Eredia. *Osservazioni meteorologiche eseguite nel 1922 nelle Stazioni istituite nella Somalia Italiana da S. A. R. il Duca degli Abruzzi*, in « Supplemento agli Annali idrografici », Vol. X, Anni 1915-1922, Genova 1924.

(2) Sul punto di licenziare le bozze del presente lavoro, troviamo (*Annali dell'Ufficio Presagi*, a cura del Prof. F. Eredia, Vol. I, 1927) essere già in uso la ricerca della risultante in riferimento ai venti spirati in un determinato luogo e periodo di tempo. Tuttavia non si tiene generalmente conto, nella determinazione di tale risultante, se non delle frequenze dei venti nelle varie direzioni della rosa dei venti: mentre l'esatta misura dell'effetto meccanico prodotto dal vento in un dato luogo e tempo può farsi soltanto considerando anche le velocità nelle stesse direzioni, come è appunto fatto nel nostro esempio.

D'altronde lo stesso concetto di risultante, considerato come un'estensione di quello di media aritmetica, si può, come abbiamo mostrato, applicare anche a serie cicliche di tutt'altro genere.

Osservazioni anemometriche eseguite alle ore 6 durante l'anno 1922
(*Villaggio S. A. R. Duca degli Abruzzi*)

Mese	Direzione	Anomalia	Somma delle velocità	Componente nella direzione *W-E*	Componente nella direzione *S-N*
Gennaio . . .	*E*	0°	3.6	3.60	0 —
	E N E	22°30′	13.9	12.84	5.32
	N E	45°	148.2	104.79	104.79
	N N E	67°30′	3.6	1.38	3.33
				122.61	113.44
Febbraio . .	*E N E*	22°30′	39.6	36.59	15.16
	N E	45°	58.2	41.15	41.15
				77.74	56.31
Marzo . . .	*N E*	45°	20.4	14.43	14.43
Aprile . . .	*S S W*	247°30′	1.2	— 0.46	— 1.11
Maggio . . .	*S W*	225°	9.6	— 6.79	— 6.79
	S S W	247°30′	2.4	— 0.92	— 2.22
				— 7.71	— 9.01
Giugno . . .	*W S W*	202′30′	9.6	— 8.87	— 3.67
	S W	225°	20.4	— 14.43	— 14.43
	S S W	247°30′	10.8	— 4.13	— 9.98
				— 27.43	— 28.08
Luglio . . .	*W S W*	202°30′	6 –	— 0.55	— 0.23
	S W	225°	15.6	— 11.03	— 11.03
				— 11.58	— 11.26
Agosto . . .	*S W*	225°	9.6	— 6.79	— 6.79
Settembre . .	*S W*	225°	6.6	— 4.67	— 4.67
Ottobre . . .	..	..	..	..	..
Novembre . .	..	..	..	..	..
Dicembre . .	*N E*	45°	28.8	16.12	16.12
	N N E	67°30′	2.4	0.92	2.22
				17.04	18.34

Omettiamo di trascrivere i prospetti analoghi per le osservazioni delle 12^h e delle 18^h.

Nel prospetto seguente abbiamo raccolto, per ciascun mese, le coppie di componenti della velocità corrispondenti a ciascuna delle ore 6, 12, 18 e abbiamo determinato per ciascuna di tali coppie il valore assoluto e l'anomalia della risultante, anomalia la quale è immediatamente riferibile alla rosa dei venti e segna in questa una direzione che, per quel certo mese e per quella certa ora di osservazione, costituisce una media nel senso meccanico stabilito.

INTENSITÀ E DIREZIO

PER CIASCUN MESE E PER CIASCUNA DELLE SERIE

(Villaggio S. A.

MESE	Osservazioni alle 6^h		
	Componenti della velocità	Risultante	
		Valore assoluto	Anomalia
Gennaio	122,61 113,44	167,04	42°46′27″
Febbraio	77,74 56,31	95,99	35°54′59″
Marzo	14,43 14,43	20,40	45°00′00″
Aprile	— 0,46 — 1,11	1,20	247°30′00″
Maggio	— 7,71 — 9,01	11,85	229°26′39″
Giugno	— 27,43 — 28,08	39,25	225°40′13″
Luglio	— 11,58 — 11,26	16,16	224°11′ 8″
Agosto	— 6,79 — 6,79	9,60	225°00′00″
Settembre	— 4,67 — 4,67	6,60	225°00′00″
Ottobre	*c*	*c*	..
Novembre	*c*	*c*	..
Dicembre	17,04 18,34	25,03	47° 6′10″

RISULTANTI DEI VENTI
AZIONI (ANTIMERIDIANE, MERIDIANE, POMERIDIANE)
egli Abruzzi, 1922)

Osservazioni alle 12^h			Osservazioni alle 18^h		
…onenti …ella …ocità	Risultante		Componenti della velocità	Risultante	
	Valore assoluto	Anomalia		Valore assoluto	Anomalia
…26,56 354,14	554,41	39°42′00″	502,70 — 29,79	503,59	356°36′31″
…26,56 354,14	433,86	28°15′21″	757,75 — 119,40	767,10	351° 2′45″
382,16 …05,39	394,15	13°43′22″	756,48 — 319,30	821,10	337° 6′56″
382,90 93,50	147,97	265° 1′26″	61,19 — 63,56	88,22	313°54′34″
12,88 …47,41	235,52	238°10′39″	5,35 — 28,44	28,94	280°39′29″
…24,19 …00,11	420,81	239°24′26″	— 54,69 — 124,83	136,28	246°20′30″
…14,16 …62,23	857,20	241°54′17″	— 253,69 — 570,46	624,33	240° 1′30″
…03,69 …56,19	709,43	245°43′10″	— 120,82 — 504,10	518,37	256′31′17″
…90,15 …94,89	661,88	243°59′58″	— 78,07 — 477,67	484,01	260°43′ 5″
13,12 74,34	75,49	79°59′35″	121,53 — 245,40	273,85	296°20′46″
…06,51 72,95	129,10	34°24′31″	261,66 — 78,75	273,25	343°14′58″
…81,31 …74,01	251,30	43°49′49″	383,84 14,74	384,13	2°11′58″

Il precedente prospetto, che sintetizza mese per mese le tre serie di osservazioni giornaliere, mette in chiara evidenza il regime dei venti durante il 1922. Per le osservazioni mattutine l'anomalia della risultante mensile varia fra 35° e 47° dal dicembre al marzo compreso, e varia invece fra 224° e 247° dall'aprile al settembre: al mattino, la regione è dunque nettamente sotto il dominio dei venti di *NE* dal dicembre al marzo, e dei venti di *SW* dall'aprile al settembre, e poichè il valore assoluto della risultante oscilla nel primo periodo da Km. 20 a Km. 167, mentre nel secondo è compreso fra Km. 1 e Km. 39, così si conclude che l'intensità del vento è, nel complesso, notevolmente superiore nel primo che nel secondo periodo.

Per le osservazioni meridiane l'anomalia della risultante mensile varia dall'ottobre al marzo con una certa continuità decrescendo da circa 80° fino a circa 14°, ossia mantenendosi nel primo quadrante; mentre dall'aprile al settembre l'anomalia stessa varia fra 238° e 265°, mantenendosi per ben cinque mesi, dal maggio al settembre, fra 238° e 246°, e cioè costantemente intorno al *SSW*. Quanto ai valori assoluti della risultante, essi variano nel primo dei detti periodi da Km. 75 a Km. 394, crescendo costantemente dall'ottobre al gennaio, e decrescendo poi fino al marzo; mentre nel secondo periodo la variazione ha luogo fra Km. 148 circa e Km. 662, crescendo costantemente fino al luglio, e poi decrescendo. Come per le osservazioni antimeridiane, così per le meridiane, l'intensità complessiva del vento è superiore nel secondo periodo a quella del primo.

Per le osservazioni pomeridiane, l'anomalia, dal novembre al marzo, oscilla fra 2° e 343° (nel tratto che comprende *O*°) cioè si mantiene intorno all'*E* e all'*ESE*; invece dall'aprile all'ottobre varia fra 240° e 296°, ossia fra il *SSW* e il *SSE*. L'intensità poi cresce continuamente nel primo periodo da Km. 273 a Km. 821, mentre nel secondo è contenuta fra il minimo di 88 Km. nell'aprile e il massimo di 624 Km. nel luglio: cioè a differenza di quanto accade nello antimeriggio e nel meriggio, la velocità complessiva è maggiore nel primo che nel secondo periodo.

b) Abbiamo fin qui considerato, per ciascuna ora di osservazione, 12 serie cicliche corrispondenti ai 12 mesi dell'anno. Ma è evidente che per studiare l'andamento dei venti in un dato luogo potrà anche convenire di sintetizzare le osservazioni giornaliere non soltanto per mesi, ma anche per stagioni, per anni o per determinati periodi. Così nell'esempio considerato sarà conveniente raccogliere,

per ciascuna ora di osservazione e in base alle constatazioni fatte sulle serie mensili, i dati rilevati dall'aprile all'ottobre (serie A), quelli dal novembre al marzo (serie B) e infine quelli per tutto il periodo annuale (serie C).

Le risultanti ottenute per ciascuna ora di osservazione addizionando le componenti mensili nella direzione W-E e nella S-N date dal primo prospetto, hanno gli elementi risultanti dall'annesso prospetto:

INTENSITÀ E DIREZIONI DELLE RISULTANTI DEI VENTI
NEI PERIODI NOVEMBRE-MARZO E APRILE-OTTOBRE PER CIASCUNA DELLE SERIE DI OSSERVAZIONI
(ANTIMERIDIANE, MERIDIANE, POMERIDIANE)
(*Villaggio S. A. R. Duca degli Abruzzi*, 1922)

PERIODO	Osservazioni alle 6^h			Osservazioni alle 12^h			Osservazioni alle 18^h		
	Componenti della velocità	Risultante		Componenti della velocità	Risultante		Componenti della velocità	Risultante	
		Valore assoluto	Anomalia		Valore assoluto	Anomalia		Valore assoluto	Anomalia
(*A*) Novembre-Marzo .	231,82 202,51	307,82	41°12′18″	1.479,45 890,00	1.731,69	31°18′48″	2.662.44 — 532.50	2.715,17	348°41′23″
(*B*) Aprile-Ottobre . .	— 58,63 — 60,91	84,54	226° 5′22″	—1.349,90 —2,781,85	3-092,07	244° 6′53″	— 319,20 —2.014,46	2 039,59	260°59′46″
(*C*) Anno 1922	173,18 141,60	223,71	39°16′16″	129,54 —1.881,85	1.886,30	273°56′20″	2.343.25 —2.546.96	3.460,16	312°36′53″

c) Delle serie (A), (B), (C) corrispondenti alle 6^h determiniamo anche le medie aritmetiche nel senso delle definizioni: $\Sigma\varepsilon = 0$ e $\Sigma\varepsilon^2 = min$. Tali serie sono:

MODALITÀ	SOMME DELLE VELOCITÀ DEL VENTO RILEVALE A 6^h NEI PERIODI		
	Novembre-Marzo (A)	Aprile-Ottobre (B)	Anno 1922 (C)
S	..	..	..
SSE	..	..	..
SE	..	..	..
ESE	..	..	..
E	3,6	..	3,6
ENE	53,5	..	53,5
NE	249,6	..	249,6
NNE	6.0	..	6,0
N	..	..	..
NNW	..	..	..
NW	..	..	..
WNW	..	..	..
W	..	..	..
WSW	..	15,6	15,6
SW	..	61,8	61,8
SSW	..	14,4	14,4
	312.7	91,8	404,5

Poichè il numero delle modalità è pari, gli intervalli determinati dai punti opposti ai punti modalità sono quelli stessi individuati dai punti modalità. Assumendo come primo intervallo S — SSE, le medie corrispondenti ai successivi intervalli hanno dal punto di

mezzo dell'intervallo $S - SSE$, preso come origine, gli scostamenti indicati dal seguente prospetto:

INTERVALLO	MEDIE NEL PERIODO		
	Novembre-Marzo (*A*)	Aprile-Ottobre (*B*)	Anno 1922 (*C*)
$S - SSE$	5,3 (*ord.*)	— 2,5	3,5 (*ord.*)
$SSE - SE$	..	..	..
$SE - ESE$	..	..	..
$ESE - E$	..	..	..
$E - ENE$	..	..	..
$ENE - NE$	..	0,2	4,1
$NE - NNE$	..	11,0	6,6
$NNE - N$	..	13,5 (*ord.*)	7,1 (*ord.*)
$N - NNW$	..	..	..
$NNW - NW$	..	..	..
$NW - WNW$	..	..	..
$WNW - W$	..	..	..
$W - WSW$	5,5	..	7,3
$WSW - SW$	8,2	..	9,4
$SW - SSW$	21,0	..	19,3
$SW - S$	21,3	..	19,5
Prova: $S - SSE$ . . .	21,3+0=5,3+16	13,5+0=−2,5+16	19,5+0=3,5+16

Per la (*A*) la sola media ordinaria è quella di scostamento 5,3 da $S - SSE$; per la (*B*) quella di scostamento 13,5; per la (*C*) esistono due medie ordinarie di scostamenti 3,5 e 7,1. È appena necessario osservare che lo scarsissimo numero di medie aritmetiche accettabili (1, 1, e 2 sopra un massimo possibile di 16 medie) dipende dalla forte asimmetria delle tre serie. Per le due prime fra queste tutte le modalità di X (direzione del vento) per cui non è nullo il corrispondente valore di Y (velocità) cadono in uno stesso emiciclo. È pertanto applicabile l'osservazione 4ª del numero 18, cioè l'unica media ordinaria della (*A*) e l'unica della (*B*) hanno comportamento

analogo alla media aritmetica di una serie rettilinea; e gli scostamenti medî, semplice e quadratico, da quelle medie coincidono con quelli che erano già stati introdotti nella tecnica statistica come scostamenti medî, semplice e quadratico, di una serie ciclica (Cfr. Oss. 2ª del n. 38 e n. 39).

d) Per le serie (*A*), (*B*), (*C*) corrispondenti alle 6^h determiniamo le mediane calcolando la somma Θ (*X*) degli scostamenti assoluti rispetto a ciascuna modalità, moltiplicati tali scostamenti per i rispettivi valori della velocità del vento; le mediane saranno, in ciascuna serie, quelle modalità per cui la detta somma risulti minima.

Per la serie (*A*) si trova:

in		*E*	*ENE*	*NE*	*NNE*	*N*	*NNW*	*NW*	*WNW*
$\Theta(X)$	=	570,7	265,2	**66,7**	367,4	680,1	992,8	1305,5	1618,2

in		*W*	*WSW*	*SW*	*SSW*	*S*	*SSE*	*SE*	*ESE*
$\Theta(X)$	=	1930,9	2236,4	2434,9	2135,0	1821,5	1508,8	1196,1	883,4

Il minimo valore di Θ (*X*) è 66,7, corrispondente alla modalità *NE*, e perciò questa costituisce la mediana della serie (*A*). Si osservi (Cfr. n. 28) che nella modalità *SW*, opposta a *NE* la $\Theta(X)$ raggiunge il suo massimo.

Similmente si determinano le mediane per le serie (*B*) e (*C*) e si trova:

Serie:	(*A*)	(*B*)	(*C*)
Mediana:	*NE*	*SW*	*NE*

Ciascuna di tali mediane insieme col punto opposto divide il ciclo in due emicicli tali che la somma delle velocità constatate nell'emiciclo precedente la mediana è minore oppure maggiore della metà della somma totale delle velocità secondochè nell'emiciclo non si comprenda oppure si comprenda la mediana stessa.

e) In *b*) abbiamo determinato le risultanti delle serie cicliche (*A*), (*B*), (*C*) corrispondenti alle 6^h; le loro anomalie rispetto alla direzione *W-E*, misurate nel senso contrario al moto degli indici di un orologio, sono rispettivamente, secondo che si prenda come unità il grado oppure lo scostamento fra due modalità consecutive:

(*A*)	$41^\circ 12' 18''$	1,831
(*B*)	$226^\circ\ 5' 22''$	10,048
(*C*)	$39^\circ 16' 16''$	1,745

Per la serie A la media aritmetica G nel senso meccanico ha dunque lo scostamento 1,831 dalla modalità E. Determinati gli scostamenti assoluti h da G delle altre modalità delle serie, si potranno di questa facilmente calcolare gli indici di variabilità ${}^1\eta_G$ e ${}^2\eta_G$ come appresso:

	Scostamento assoluto da G h	h^2	Velocità v	hv	$h^2 v$
E	1,831	3,352561	3,6	6,592	12,069 220
ENE	0,831	0,690561	53,5	44,459	36,945 014
NE	0,169	0,028561	249,6	42,182	7,128 826
NNE	1,169	1,366561	6,0	7,014	8,199 366
			$\Sigma v = 312,7$	$\Sigma hv = 100,247$	$\Sigma h^2 v = 64,342426$

$${}^1\eta_G = \frac{\Sigma\, h v}{\Sigma\, v} = \frac{100,247}{312,7} = 0,321$$

$${}^2\eta_G = \left(\frac{\Sigma\, h^2 v}{\Sigma\, v}\right)^{\frac{1}{2}} = \left(\frac{64,342426}{312,7}\right)^{\frac{1}{2}} = 0,454$$

Similmente si procederà per le altre due serie, e si troverà:

	A	B	C
${}^1\eta_G$	0,321	0,360	1,993
${}^2\eta_G$	0,454	0,572	3,603

Si constata che le serie (A) e (B) danno luogo per l'indice di mutabilità ${}^1\eta_G$ ed anche per l'indice ${}^2\eta_G$ a valori pressochè uguali, mentre la serie (C) dà luogo a corrispondenti valori notevolmente superiori ai primi: come per le serie rettilinee, così per le cicliche accade che per una serie (C), risultante dall'insieme di altre due (A) e (B), un certo indice di mutabilità acquista generalmente valore superiore a quelli acquistati per queste singole serie.

Osservazioni sui mesi di nascita delle madri e dei primogeniti in Matelica e in Roma

60. — *a*) Su 1733 osservazioni fatte in Matelica e su 4143 fatte in Roma, dal Boldrini (1), il mese di nascita della madre e quello del primogenito risultarono come appresso:

	Matelica		Roma	
	Per le madri	Per i primogeniti	Per le madri	Per i primogeniti
Gennaio	154	167	401	369
Febbraio	178	162	391	329
Marzo	224	168	417	384
Aprile	218	168	307	329
Maggio	176	128	333	310
Giugno	121	150	291	318
Luglio	107	109	324	377
Agosto	100	127	333	402
Settembre	108	157	339	357
Ottobre	106	138	310	302
Novembre	122	132	322	307
Dicembre	119	127	375	359
	1733	1733	4143	4143

Determiniamo anzitutto le medie aritmetiche di ciascuna delle quattro serie cicliche, a sensi delle definizioni basate sulle proprietà $\Sigma\varepsilon = o$, $\Sigma\varepsilon^2 = min.$ Si debbono considerare gli intervalli determinati dai punti opposti a tutti i 12 punti modalità, dato che i corrispondenti valori di Y sono tutti diversi da zero. Se diciamo x_1 la media corrispondente all'intervallo *Lug'*...*Ag'*, coincidente con *Genn*...*Febb*, e percorriamo poi il ciclo nel senso positivo determinato dall'ordine consueto di enunciazione dei mesi dell'anno, troveremo come sco-

(1) V. G. Pietra, *The Theory of statistical relations*, ecc. già cit., in «Metron» I - VI - 1925, p. 545.

stamento delle medie successive dal punto di mezzo del primo intervallo *Genn...Febb*:

	Matelica		Roma	
	Per le madri	Per i primogeniti	Per le madri	Per i primogeniti
x_{I}	0.433	—0,010	—0,040	—0,051
x_{II}	1.126	0.870	0,925	1,113
x_{III}	1,874	1.957	1,907	2,147
x_{IV}	2,608	2,913	2,805	3,022
x_{V}	3.453	3,827	3,737	3.911
x_{VI}	4,277	4.706	4.823	4.951
x_{VII}	5.343	5.863	5,985	6.020
x_{VIII}	6,576	6,984	7,711	6,973
x_{IX}	8,127	8.147	8,325	8,085
x_{X}	9.636	9.311	9,214	9,038
x_{XI}	10,855	10,097	10,179	9,936
x_{XII}	11,693	11,235	11.022	10,857

mentre gli scostamenti delle stesse medie dai punti di mezzo degli intervalli corrispondenti saranno:

	Matelica		Roma	
	Per le madri	Per i primogeniti	Per le madri	Per i primogeniti
x_{I}	0,433	—0,010	—0,040	—0,051
x_{II}	0,126	—0,130	—0,075	0,113
x_{III}	—0,126	—0,043	—0,093	0,147
x_{IV}	—0.392	—0,087	—0.195	0,022
x_{V}	—0,547	—0.173	—0,263	—0,089
x_{VI}	—0,723	—0,294	—0,177	—0,049
x_{VII}	—0,657	—0,137	—0,015	0,020
x_{VIII}	—0,424	—0,016	0,117	—0,027
x_{IX}	0,127	0,147	0.325	0,085
x_{X}	0,636	0,311	0,214	0,038
x_{XI}	0,855	0,097	0,179	—0,064
x_{XII}	0,693	0,236	0,022	—0,143

e perciò le medie aritmetiche accettabili (tutte ordinarie) sono:

MATELICA		ROMA	
Per le madri	Per i primogeniti	Per le madri	Per i primogeniti
x_{I}	x_{I}	x_{I}	x_{I}
x_{II}	x_{II}	x_{II}	x_{II}
x_{III}	x_{III}	x_{III}	x_{III}
x_{IV}	x_{IV}	x_{IV}	x_{IV}
. .	x_{V}	x_{V}	x_{V}
. .	x_{VI}	x_{VI}	x_{VI}
. .	x_{VII}	x_{VII}	x_{VII}
x_{VIII}	x_{VIII}	x_{VIII}	x_{VIII}
x_{IX}	x_{IX}	x_{IX}	x_{IX}
. .	x_{X}	x_{X}	x_{X}
. .	x_{XI}	x_{XI}	x_{XI}
. .	x_{XII}	x_{XII}	x_{XII}

Per Matelica, il fatto che la prima serie ammette soltanto 6 medie ordinarie e la seconda 12, denuncia nelle serie delle nascite dei primogeniti secondo i mesi una maggiore simmetria che nella serie analoga relativa alle nascite delle madri. Per Roma, invece, non è applicabile questo criterio differenziale, presentando le due serie di nascite lo stesso numero, ed anzi il massimo possibile, di medie aritmetiche ordinarie. Il paragone fra la simmetria dell'una e quella dell'altra richiederebbe, come altrove si accennò, una analisi più minuta.

b) Determiniamo anche, per le stesse serie, le medie aritmetiche in senso meccanico, in base alla considerazione dei vettori aventi come direzioni quelle determinate dai punti rappresentativi dei mesi, e come moduli i rispettivi numeri di nascite. La scomposizione di ciascun vettore si farà lungo gli assi ortogonali ξ e η, rispettivamente determinati dai punti immagini del gennaio e dello aprile.

Per le serie delle nascite in Matelica si avrà:

	Numero delle nascite		Componente su ξ		Componente su η	
	Madri	Primogeniti	Madri	Primogeniti	Madri	Primogeniti
Gennaio . . .	154	167	154	167	0	0
Febbraio . . .	178	162	154,15	140,30	89	81
Marzo	224	168	112	84	193,99	145,49
Aprile	218	168	0	0	218	168
Maggio . . .	176	128	— 88	— 64	152,42	110,85
Giugno . . .	121	150	— 104,79	— 129,90	60,50	75
Luglio	107	109	— 107	— 109	0	0
Agosto	100	127	— 86,60	— 109,99	— 50	— 63,50
Settembre . .	108	157	— 54	— 78,50	— 93,53	— 135,97
Ottobre . . .	106	138	0	0	— 106	— 138
Novembre . .	122	132	61	66	— 105,66	— 114,32
Dicembre . .	119	127	103,06	109,99	— 59,50	— 63,50
	1733	1733	143,82	75,90	299,22	65,05

Ne segue che per le nascite delle madri la risultante ha come modulo: $R_m = 331,99$ e come anomalia $\varphi_m = 64^\circ 40' 16''$ mentre per quelle dei primogeniti $R_f = 99,96$ e $\varphi_f = 40^\circ 37' 38''$.

Il fatto che $R_f < R_m$, dato che i moduli dei vettori dei due sistemi hanno somme uguali, può anche esso interpretarsi come denotante una maggiore simmetria nelle nascite dei primogeniti che non nelle nascite delle madri. La media rappresentata dall'anomalia $\varphi_m = 64^\circ 40' 16''$ denota una modalità del carattere « mese

dell'anno » compresa fra il marzo e l'aprile; e quella rappresentata da $\varphi_f = 40^\circ 37' 38''$ una modalità compresa fra il febbraio ed il marzo.

In modo del tutto analogo si troverebbe per le serie relative a Roma:

come modulo $R_m = 241{,}33$ e come anomalia $\varphi_m = 14^\circ 40' 26''$ (per le madri);

come modulo $R_f = 24{,}05$ e come anomalia $\varphi_f = 189^\circ 37' 17''$ (per i primogeniti).

Anche qui il confronto dei moduli delle due risultanti rivela una maggiore simmetria nelle nascite dei primogeniti che nelle nascite delle madri, benchè nessuna conclusione al riguardo fosse possibile trarre dal confronto dei numeri delle medie aritmetiche nel senso delle definizioni *c*) e *d*). Le anomalie delle risultanti, le quali individuano le medie nel senso della definizione *e*), corrispondono a modalità quasi opposte del ciclo, e quindi anche più lontane di quanto era risultato per le due serie di Matelica. Che cosa può significare, *in questi esempi*, la diminuzione di valore assoluto della risultante passando dalle nascite delle madri a quelle dei primogeniti, e il notevole distacco delle due anomalie ? Riflettiamo che se i primogeniti ereditassero perfettamente il mese di nascita materno le due risultanti, per le nascite delle madri e per quelle dei primogeniti, coinciderebbero (in valore assoluto e anomalia). La mancata coincidenza può attestare che tale ereditarietà non si verifica; ma di più il diverso valore assoluto e la diversa anomalia delle risultanti possono anche, in un certo modo, misurare il grado di non ereditarietà del mese di nascita. Mentre la prima di tali circostanze differenziali può interpretarsi come corrispondente a un diverso grado di simmetria nelle due distribuzioni di nascite, la seconda sembra prevalentemente significare che tanto maggiore è la divergenza delle anomalie tanto minore è la concordanza fra le due serie di nascite. Il secondo degli esempi, cioè quello di Roma, in cui le anomalie denotano modalità quasi opposte del ciclo annuale, mette specialmente in luce tale mancata concordanza (1).

(1) Il BOLDRINI, *L'epoca di generazione.* « Rivista di Antropologia», 1919, ed il PIETRA (op. cit.), hanno per via diversa e in base a copioso materiale statistico, dimostrato la mancanza di qualsiasi connessione tra il mese di nascita della madre e quello del primogenito. Sullo stesso argomento vedi anche: L. GALVANI. *La stagionalità delle nascite nelle singole famiglie,* « Metron », Vol. VI.

Morti della città di Roma, per mesi e per classi di età, nel triennio 1925-27 (1)

61. — Un esempio di serie statistica dipendente da una variabile e da una mutabile ciclica, è fornito dal seguente prospetto, in cui i morti della Città di Roma nel triennio 1925-27 sono stati distribuiti a seconda dell'età e a seconda del mese in cui avvenne la morte.

(1) I morti in età ignota (21 in tutto) sono stati distribuiti fra le varie classi in proporzione al loro ammontare.

MORTI PER CLASSI DI ETÀ E PER MESI (1)
(Roma, 1925-27)

CLASSI DI ETÀ	ETÀ CENTRALE 1 = anni 10	GENNAIO	FEBBRAIO	MARZO	APRILE	MAGGIO	GIUGNO	LUGLIO	AGOSTO	SETTEMBRE	OTTOBRE	NOVEMBRE	DICEMBRE	$L_i =$
0	0,05	423	342	362	319	345	593	536	412	266	306	276	405	4.585
1	0,15	182	229	261	237	186	270	353	277	180	192	183	191	2.741
2	0,25	92	113	110	122	98	90	89	76	61	51	62	81	1.045
3	0,35	36	48	75	59	52	55	44	36	33	23	35	51	547
4	0,45	37	34	36	31	29	36	34	29	17	25	20	34	362
5-9	0,75	53	54	55	72	68	55	57	49	45	46	62	52	668
10–14 . . .	1,25	44	50	54	54	47	46	47	45	40	38	46	43	554
15–19 . . .	1,75	96	94	118	80	103	105	115	89	83	104	105	112	1.204
20–24 . . .	2,25	130	136	141	148	125	141	138	136	131	136	124	130	1.616
25–29 . . .	2,75	112	118	123	129	98	113	107	118	103	101	112	116	1.350
30–34 . . .	3,25	96	108	113	89	98	86	91	77	87	111	94	112	1.162
35–39 . . .	3,75	109	104	125	108	94	95	92	100	89	108	89	107	1.220
40–44 . . .	4,25	130	128	130	101	96	107	88	89	84	95	81	135	1.264
45–49 . . .	4,75	116	121	126	115	99	106	116	124	80	107	108	111	1.329
50–54 . . .	5,25	148	142	169	143	135	116	104	111	126	126	123	181	1.624
55–59 . . .	5,75	187	205	194	173	157	140	161	115	147	157	164	217	2.017
60–64 . . .	6,25	281	273	267	191	183	178	185	186	163	195	192	286	2.580
65–69 . . .	6,75	320	301	318	210	239	195	203	186	135	186	231	313	2.837
70–79 . . .	7,50	654	608	545	405	359	353	337	320	296	357	393	667	5.294
80 e più. .	8,50	302	295	272	197	157	143	148	127	131	182	195	303	2.452
	$C_j =$	3.548	3.503	3.594	2.983	2.768	3.023	3.045	2.702	2.297	2.646	2.695	3.647	36.451

(1) Dati cortesemente forniti dal Prof. L. MAROI, Direttore della IX Ripartizione del Governatorato di Roma.

a) Si è detto che, in questo caso, la sola definizione di media aritmetica che si possa applicare è quella di modalità rispetto alla quale sia minimo il momento secondo; e si è aggiunto che la definizione stessa può avere significato sia nel solo ambito delle modalità effettive della serie, sia in quello più ampio ottenuto con l'introduzione di modalità fittizie. Ci proponiamo, appunto, di determinare la media aritmetica in senso ristretto, cioè *limitandoci a considerare le sole modalità effettive* della serie; e dovremo, pertanto, calcolare il momento secondo della serie rispetto a ciascuna coppia di modalità. Per determinare gli scostamenti fra le varie coppie di modalità età-mese e in particolare quelli fra due classi di età sarebbe necessario sostituire a ciascuna classe la rispettiva età media: ma supponendo, in prima approssimazione, che entro ciascuna di tali classi la distribuzione delle età sia uniforme (Cfr. nota al n. 7), sostituiremo ad ogni classe la rispettiva età centrale; alla classe aperta « 80 e oltre » sostituiamo poi l'età 85. Inoltre, perchè gli scostamenti fra le modalità del carattere « età » siano di un ordine di grandezza paragonabile con quelli del carattere « mese », prendiamo come unità di tempo il decennio anzichè l'anno, cosicchè lo scostamento massimo fra le età risulterà 8,5 — 0,05 = 8,45 e quello massimo fra i mesi sarà di 6 unità. I valori delle età impiegati per il calcolo degli scostamenti sono indicati nella seconda colonna del prospetto fondamentale. Con tali semplificazioni si potrà dunque dire che lo scostamento fra le classi 1 e 5-9 è 0,75 — 0,05 = 0,70 unità, quella fra le classi 10-14 e 65-69 è 6,75 — 1,25 = 5,50, ecc.

Ciò posto, detti $d_{(e,\, m)\,(i,\, j)}$, $d_{(e,\, i)}$, $d_{(m,\, j)}$ gli scostamenti dalla coppia (età e, mese m) alla coppia (età i, mese j), dalla età e alla età i, dal mese m al mese j, si avrà:

$$d^2_{(e,m)\,(i,j)} = d^2_{(e,i)} + d^2_{(m,j)}$$

e perciò il momento secondo della serie rispetto alla coppia di modalità $(e,\, m)$ sarà:

$$\mu_2\,(e,\, m) = \sum_i \sum_j d^2_{(e,i)}\, T_{i\,j} + \sum_i \sum_j d^2_{(m,j)}\, T_{i\,j}$$

essendo $T_{i,j}$ il numero dei morti di età i nel mese j.

Pertanto:

$$\begin{aligned} \mu_2\,(e,m) &= \sum_i d^2_{(e,i)} \sum_j T_{j\,i} + \sum_j d^2_{(m,j)} \sum_i T_{i\,j} \\ &= \sum_i d^2_{(e,i)}\, L_i + \sum_j d^2_{(m,j)} \cdot C_j \end{aligned}$$

dove L_i e C_j denotano rispettivamente le somme dei morti per linee (età i) e per colonne (mese j).

Infine, ponendo

$$A_e = \sum_i d^2_{(e,i)} L_i \,, \quad B_m = \sum_j d^2_{(m,j)} C_j$$

risulterà

$$\mu_2 (e, m) = A_e + B_m \,,$$

dove A_e è il momento secondo del sistema di valori L_i rispetto alla età e e B_m è il momento secondo del sistema C_j rispetto al mese m.

Si ha poi:

$$A_e = \sum_i d^2_{(e,i)} L_i = \Sigma (i - e)^2 L_i = e^2 \sum_i L_i - 2 e \sum_i i L_i + \sum_i i^2 L_i$$
$$= \alpha e^2 - 2 \beta e + \gamma \,.$$

cosicchè, calcolati i coefficienti α, β, γ, i valori A_e si avranno senz'altro come ordinate di una parabola. Nel nostro caso il calcolo dei momenti A_e risulta dai due seguenti prospetti; nel secondo dei quali si può subito osservare che il minimo valore di A_e si verifica per la età 42,5 ed è 323.638,1350

Calcolo dei coefficienti della parabola $\alpha e^2 - 2\beta e + \gamma$

Età i	L_i	$i L_i$	$i^2 L_i$
0,05	4.585	229,25	11,4625
0,15	2.741	411,15	61,6725
0,25	1.045	261,25	65,3125
0,35	547	191,45	67,0075
0,45	362	162,90	73,3050
0,75	668	501,00	375,7500
1,25	554	692,50	865,6250
1,75	1.204	2.107,00	3.687,2500
2,25	1.616	3.636,00	8.181,0000
2,75	1.350	3.712,50	10.209,3750
3,25	1.162	3.776,50	12.273,6250
3,75	1.220	4.575,00	17.156,2500
4,25	1.264	5.372,00	22.831,0000
4,75	1.329	6.312,75	29.958,5625
5,25	1.624	8.526,00	44.761,5000
5,75	2.017	11.597,75	66.687,0625
6,25	2.580	16.125,00	100.781,2500
6,75	2.837	19.149,75	129.260,8125
7,50	5.294	39.705,00	297.787,5000
8,50	2.452	20.842,00	177.157,0000
	$\alpha = 36.451$	$\beta = 147.886,75$	$\gamma = 922.279,3225$

CALCOLO DEI MOMENTI DI SECONDO ORDINE A_e

ETÀ e 1	e^2 2	$\alpha\ e^2$ 3	$-2\beta e+\gamma$ 4	A_e 3 + 4
0,05 . . .	0,0025	91,1275	907.490,6475	907.581,7750
0,15 . . .	0,0225	820,1475	877.913,2975	878.733,4450
0,25 . . .	0,0625	2.278,1875	848.335,9475	850.614,1350
0,35 . . .	0,1225	4.465.2475	818.758,5975	823.223,8450
0,45 . . .	0,2025	7.381,3275	789.181,2475	796.562,5750
0,75 . . .	0,5625	20.503,6875	700.449,1975	720.952,8850
1,25 . . .	1,5625	56.954,6875	552.562,4475	609.517,1350
1,75 . . .	3,0625	111.631,1875	404.675,6975	516.306,8850
2,25 . . .	5,0625	184.533,1875	256.788,9475	441.322,1350
2,75 . . .	7,5625	275.660,6875	108.902,1975	384.562,8850
3,25 . . .	10,5625	385.013,6875	— 38.984,5525	346.029,1350
3,75 . . .	14,0625	512.592,1875	— 186.871,3025	325.720,8850
4,25 . . .	18,0625	658.396,1875	— 334.758,0525	**323.638,1350**
4,75 . . .	22,5625	822.425,6875	— 482.644,8025	339.780,8850
5,25 . . .	27,5625	1.004.680,6875	— 630.531,5525	374.149,1350
5,75 . . .	33,0625	1.205.161,1875	— 778.418,3025	426.742,8850
6.25 . . .	39,0625	1.423.867,1875	— 926.305,0525	497.562,1360
6,75 . . .	45,5625	1.660.798,6875	— 1.074.191,8025	586.606,8850
7,50 . . .	56,2500	2.050.368,7500	— 1.296.021,9275	754.346,8225
8,50 . . .	72,2500	2.633.584,7500	— 1.591.795,4275	1.041.789,3225

Quanto ai momenti secondi B_m, basta osservare che gli scostamenti $d_{(m,j)}$ fra i mesi assumono sempre, in ordine ciclico e rispetto a qualunque modalità, gli stessi valori

$$0,\quad 1,\quad 2,\quad 3,\quad 4,\quad 5,\quad 6,\quad 5,\quad 4,\quad 3,\quad 2,\quad 1,$$

cosicchè sarà facile, come risulta dal seguente prospetto, dedurre dai valori C_j e dai quadrati di detti scostamenti i valori B_m.

MESI	MORTI C_j	GENNAIO	FEBBRAIO	MARZO	APRILE
Gennaio	3.548	..	3.548	14.192	31.932
Febbraio	3.503	3.503	..	3.503	14.012
Marzo	3.594	14.376	3.594	..	3.594
Aprile	2.983	26.847	11.932	2.983	..
Maggio	2.768	44.288	24.912	11.072	2.768
Giugno	3.023	75.575	48.368	27.207	12.092
Luglio	3.045	109.620	76.125	48.720	27.405
Agosto	2.702	67.550	97.272	67.550	43.232
Settembre	2.297	36.752	57.425	82.692	57.425
Ottobre	2.646	23.814	42.336	66.150	95.256
Novembre	2.695	10.780	24.255	43.120	67.375
Dicembre	3.647	3.647	14.588	32.823	58.352
Somme B_m	36.451	416.752	404.355	400.012	413.443

Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre
56.768	88.700	127.728	88.700	56.768	31.932	14.192	3.548
31.527	56.048	87.575	126.108	87.575	56.048	31.527	14.012
14.376	32.346	57.504	89.850	129.384	89.850	57.504	32.346
2.983	11.932	26.847	47.728	74.575	107.388	74.575	47.728
..	2.768	11.072	24.912	44.288	69.200	99.648	69.200
3.023	..	3.023	12.092	27.207	48.368	75.575	108.828
12.180	3.045	..	3.045	12.180	27.405	48.720	76.125
24.318	10.808	2.702	..	2.702	10.808	24.318	43.232
36.752	20.673	9.188	2.297	..	2.297	9.188	20.673
66.150	42.336	23.814	10.584	2.646	..	2.646	10.584
97.020	67.375	43.120	24.255	10.780	2.695	..	2.695
91.175	131.292	91.175	58.352	32.823	14.588	3.647	..
436.272	467.323	483.748	487.293	480.928	460.579	441.540	428.971

Si noti anche qui che il minimo valore di B_m si verifica nel mese di marzo ed è 400.012.

Si conclude che la somma dei due momenti secondi A_e e E_m cioè $\mu_2\,(e,\,m)$ assume il minimo valore 723.650,14 per la coppia di modalità (età in decenni: 4,25; mese: marzo) la quale coppia costituisce dunque la media aritmetica della serie, nel senso inizialmente dichiarato, cioè quando ci si limiti a considerare le sole modalità effettive della serie. Si conclude altresì che lo scostamento quadratico medio da tale media è

$$ {}^{2}\eta_{A} = \left(\frac{723650{,}14}{36.451}\right)^{\frac{1}{2}} = 4{,}46\,; $$

mentre il valore dello scostamento semplice medio dalla media stessa si troverebbe calcolando la somma Θ degli scostamenti assoluti dalla coppia di modalità costituente la media aritmetica, e sarebbe pertanto:

$$ {}^{1}\eta_{A} = \frac{\Theta\,(4{,}25 - \text{marzo})}{36.451} = \frac{151.488}{36.451} = 4{,}16 $$

Si osservi che il calcolo della media o delle medie aritmetiche della data serie in senso generale, cioè comprendendo oltre le modalità effettive anche quelle di conto, sarebbe molto più laborioso. Una maggiore precisione si potrebbe conseguire dividendo gli intervalli di variazione dei due caratteri in un maggior numero di modalità effettive: così per le età si potrebbero considerare i singoli anni, e il ciclo annuale si potrebbe dividere in 24 mezzi mesi. Comunque, al fine di determinare la tendenza complessiva di serie congeneri alla data verso una particolare coppia di caratteri (media aritmetica) potrà essere sufficiente la distinzione in modalità da noi adottata. Sarebbe interessante, attraverso molte serie analoghe a quella presa in esame, constatare come tenda a spostarsi la coppia media di modalità da tempo a tempo e da luogo a luogo.

Regime dei venti osservati a De Bilt nel biennio 1925-26

62. — Dalla serie delle osservazioni anemometriche orarie (direzione e velocità del vento) fatte a De Bilt (1), durante il biennio 1925-

(1) *Annuaire* del « Koninklijk Nederlandsch Meteorologisch Instituut », *A. Météorologie*, 1925; ibid. 1926.

1926, abbiamo ricavato il prospetto seguente nel quale, internamente ad ogni casella, è inscritta la somma delle velocità del vento osservate nel biennio stesso in corrispondenza alla direzione denotata dalla linea e al mese denotato dalla colonna a cui la casella stessa appartiene. Evidentemente si ha, dal prospetto costruito, l'esempio di una serie statistica dipendente da due mutabili cicliche: X = direzione del vento e Y = mese dell'anno; il carattere quantitativo funzione di tali mutabili è T = somma delle velocità del vento. Brevemente: $T = S\,((X, Y))$.

Ci proponiamo di determinare la media aritmetica di tale serie in base all'unica definizione applicabile in questo caso (Cfr. n. 55) cioè a quella secondo cui viene riguardata come media aritmetica la coppia di modalità X, Y rispetto alla quale sia minima la somma dei quadrati degli scostamenti delle altre coppie di modalità intendendo che tali quadrati siano ponderati mediante i valori di T corrispondenti a queste coppie. Ora, analogamente a quanto si è osservato per le serie cicliche (dipendenti da una sola mutabile) (Cfr. n. 21) la posta definizione può avere significato sia in riferimento alle sole coppie di modalità effettivamente date per X ed Y, sia nell'insieme più ampio di coppie-modalità che si otterrebbe introducendo fra le modalità effettive di X e di Y altre modalità di conto. Così nel nostro esempio le coppie-modalità da considerare nel primo significato sono le 16×12, che risultano dal combinare ciascuna delle 16 direzioni della rosa dei venti con ciascuno dei 12 mesi dell'anno; nell'altro significato si dovrebbero anche considerare le combinazioni delle modalità di conto di X con le modalità effettive e di conto di Y. Rappresentando geometricamente le coppie di modalità di X ed Y mediante punti della superficie di un toro, si dovranno considerare nel primo significato soltanto i 16×12 punti intersezioni dei meridiani che rappresentano le modalità effettive di una mutabile coi paralleli che rappresentano le modalità effettive dell'altra mutabile; mentre nel secondo significato si dovranno considerare anche gli altri punti della stessa superficie ottenuti come intersezioni di meridiani e di paralleli diversi da quei primi. Se poi si ricorre alla rappresentazione sopra un piano, come è suggerito al numero 55, allora basterà considerare la configurazione costituita dai centri delle 16×12 caselle del prospetto costruito per avere l'immagine delle coppie-modalità effettive del nostro esempio, mentre i punti diversi da tali centri sarebbero immagini di coppie-modalità di cui una, almeno, non effettiva.

De Bilt. — SOMME DELLE VELOCITÀ DEL VENTO

DIREZIONI / MESI	GENNAIO	FEBBRAIO	MARZO	APRILE	MAGGIO
E	341,1	260,7	386,7	216,2	191,2
E N E	292,0	225,9	426,4	153,5	214,8
N E	42,9	195,2	719,9	436,8	759,6
N N E	30,5	38,5	625,6	495,7	430,6
N	24,2	77,9	505,9	183,3	389,3
N N W	44,5	71,0	341,6	254,8	284,5
N W	143,7	134,2	643,7	291,3	545,0
W N W	292,8	444,9	1.274,8	419,8	330,0
W	571,7	672,6	1.317,3	469,0	279,8
W S W	769,3	627,6	445,4	508,9	233,0
S W	1.698,5	1.100,4	722,3	714,8	899,2
S S W	1.636,6	1.557,9	330,0	784,9	1.064,9
S	748,6	1.150,6	112,5	557,4	632,9
S S E	572,6	487,7	82,9	307,5	317,1
S E	536,2	546,6	108,2	483,2	556,7
E S E	372,7	269,8	124,4	347,5	262,0
$C_j =$	8.117,9	7.861,5	8.167,6	6.624,6	7.390,6

OSSERVATE NELLE VARIE DIREZIONI, NEL BIENNIO 1925-1926

Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre	$Li =$
200,4	375,8	88,5	12,5	104,3	176,2	74,7	2.428,3
109,9	213,3	44,8	13,4	287,6	548,0	325,0	2.854,6
291,8	521,5	68,6	102,5	586,5	309,4	701,8	4.736,5
310,0	640,9	215,3	127,3	397,4	151,1	48,3	3.511,2
497,1	286,9	229,9	219,3	187,5	186,3	110,8	2.898,4
695,8	481,1	309,5	278,0	222,6	137,9	285,0	3.406,3
1.083,9	481,8	526,2	441,8	278,1	166,4	495,8	5.231,9
598,6	655,7	684,9	643,5	462,1	149,6	618,9	6.575,6
437,6	423,9	621,9	826,0	670,5	228,3	1.237,0	7.755,6
173,0	425,2	444,3	555,2	364,7	317,0	1.236,5	6.100,1
433,1	727,2	759,0	921,1	789,6	770,7	1.296,4	10.832,3
420,2	545,4	766,1	856,8	1.173,6	1.005,2	621,0	10.762,6
132,0	231,8	278,5	525,9	491,5	1.363,9	345,2	6.570,8
127,6	103,5	148,5	248,9	248,6	810,9	524,0	3.979,8
235,2	172.8	319,8	359,4	475,4	748,7	235,7	4.777,9
141,9	230,9	184,4	84,9	182,2	224,9	107,2	2.532,8
5.888,1	6.517,7	5.690,2	6.216,5	6.922,2	7.294,5	8.263,3	84.954,7

Analogamente a quanto si fece nell'esempio precedente *ci limiteremo a considerare le sole coppie di modalità effettive, e a ricercare quella o quelle fra tali coppie che soddisfano alla ricordata definizione di media aritmetica*, ossia quella o quelle coppie rispetto a cui risulti minimo il momento secondo della serie data.

Se indichiamo con $\mu_2\ (l, c)$ il momento secondo della serie rispetto alla coppia di modalità rappresentata nel nostro quadro dalla casella che occupa la linea l e la colonna c, con $d_{(e,c)\,(i,j)}$ lo scostamento da tale coppia ad un'altra coppia generica corrispondente alla linea i e alla colonna j, con $d_{(e,i)}$ e $d_{(c,j)}$ gli scostamenti dalla linea l alla i, e dalla colonna c alla j, si avrà evidentemente

$$d^2_{(l,c)\,(i,j)} = d^2_{l,\,i} + d^2_{c,\,i}$$

e perciò

$$\begin{aligned} \mu_2\ (l,c) &= \sum_i \sum_j d^2_{(l,c)\,(i,j)}\, T_{i\,j} = \sum_i \sum_j (d^2_{l,i} + d^2_{c,j})\, T_{i\,j} = \\ &= \sum_i d^2_{(l,\,i)} \sum_j T_{i\,j} + \sum_j d^2_{(c,\,j)} \sum_i T_{i\,j} \end{aligned}$$

ossia, indicando con L_i e C_j le somme dei valori di T (nel nostro caso velocità del vento) rispettivamente nella linea i e nella colonna j:

$$\mu_2\ (l,c) = \Sigma\, d^2_{(l,i)}\, L_i + \Sigma\, d^2_{(c,j)}\, C_j = A_l + B_c\ ,$$

dove A_l e B_c hanno evidentemente significato di momenti secondi parziali.

Si deve, naturalmente, ricordare che lo scostamento dalla linea fissa l alla linea variabile i, e similmente quello dalla colonna c alla j, non sono altro che scostamenti di modalità di due caratteri ciclici; perciò, nel nostro caso, presentando il carattere « direzione del vento », rappresentato nelle varie linee, 16 modalità e il carattere « mese dell'anno », rappresentato nelle varie colonne, 12 modalità, lo scostamento $d_{(l,i)}$ avrà in ordine circolare i valori:

0 1 2 3 4 5 6 7 8 7 6 5 4 3 2 1

e lo scostamento $d_{(c,j)}$ i valori:

0 1 2 3 4 5 6 5 4 3 2 1

Tenuto conto di ciò, si sono costruiti separatamente due prospetti, uno per il calcolo dei momenti parziali A_l, l'altro per quello dei momenti B_c. Per esempio, nel primo di essi la prima colonna è così formata, $0.L_1 = 0$; $1.L_2 = 2854{,}6$; $4.L_3 = 18.946{,}0$; $9.L_4 =$ $= 31{,}600{,}8$ .... $9.L_{14} = 35.818{,}2$; $4.L_{15} = 19.111{,}6$; $1.L_{16} =$ $= 2.532{,}8$.

Infine, eseguendo tutte le possibili somme della forma $A_l + B_c$, si può avere in corrispondenza a ciascuna coppia di modalità effettive il momento secondo della data serie. È ovvio poi che questo assume il minimo valore per la coppia costituita dal minimo valore di A_l e dal minimo di B_c, cioè nel nostro caso, per la coppia di modalità *Febbraio* — *W SW*, la quale pertanto costituisce, nel senso dichiarato, la media aritmetica della data serie.

Dal fatto poi che per tale coppia di modalità *Febbraio* — *WSW* il momento secondo della serie è $\mu_2 = min\ A_l + min\ B_c = 2.191.773{,}7$ si deduce subito che lo scostamento quadratico medio dalla media è:

$$^2\eta_A = \left(\frac{2.191.773{,}7}{84.954{,}7}\right)^{\frac{1}{2}} = 5{,}08$$

Il calcolo dello scostamento semplice medio dalla media $^1\eta_A$ riesce, invece, alquanto laborioso, per effetto della irrazionalità degli scostamenti fra la coppia di modalità costituente la media aritmetica e le altre coppie di modalità. La somma degli scostamenti assoluti della serie rispetto ad una coppia generica di modalità (l, c) sarà infatti, con le stesse notazioni usate poc'anzi:

$$\Theta(l, c) = \sum_i \sum_j d_{(l,c)\,(i,j)}\, T_{i\,j}$$

essendo

$$d_{(l,c)\,(i,j)} = \left| \left\{ d^2_{(l,i)} + d^2_{(c,j)} \right\}^{\frac{1}{2}} \right|$$

e intendendo che le $d_{(l,i)}$ denotino gli scostamenti fra le modalità del carattere ciclico rappresentato dalle linee e similmente le $d_{(c,j)}$ siano gli scostamenti fra le modalità del carattere ciclico rappresentato dalle colonne del dato prospetto.

Nel nostro esempio, lo specchio degli scostamenti $d_{(l,c)\,(i,j)}$ delle varie coppie di modalità da una di esse, p. es. dalla modalità *E — Genn.* per cui $l = 1$, $c = 1$, è dato successivamente ai prospetti dei momenti secondi parziali A_l e B_c.

CALCOLO DELLE A_l, CIOÈ DELLE SOMME DEI QUADRATI DEGLI SCOSTA

	E	*E N E*	*N E*
E	. .	2.428,3	9.713,2
E N E	2.854,6	. .	2.854,6
N E	18.946,0	4.736,5	. .
N N E	31.600,8	14.044,8	3.511,2
N	46.374,4	26.085,6	11.593,6
N N W	85.157,5	54.500,8	30.656,7
N W	188.348,4	130.797,5	83.710,4
W N W	322.204,4	236.721,6	164.390,0
W	496.358,4	380.024,4	279.201.6
W S W	298.904,9	390.406,4	298.904,9
S W	389.962,8	530.782,7	693.267,2
S S W	269.065,0	387.453,6	527.367,4
S	105.132,8	164.270,0	236.548,8
S S E	35.818,2	63.676,8	99.495,0
S E	19.111,6	43.001,1	76.446,4
E S E	2.532,8	10.131,2	22.795,2
A_l = . . .	2.312.372,6	2.439.061,3	2.540.456,2

ENTI RISPETTO ALLE DIREZIONI – (*De Bilt*, Velocità del vento, 1925-1926).

N N E	N	N N W	N W	W N W
21.854,7	38.852,8	60.707,5	87.418,8	118.986,7
11.418,4	25.691,4	45.673,6	71.365,0	102.765,6
4.736,5	18.946,0	42.628,5	75.784,0	118,412,5
..	3.511,2	14.044,8	31.600,8	56.179,2
2.898,4	..	2.898,4	11.593,6	26.085,6
13.625,2	3.406,3	..	3.406,3	13.625,2
47.087,1	20.927,6	5.231,9	..	5.231,9
105,209,6	59.180,4	26.302,4	6.575,6	..
193.890,0	124.089,6	69.800,4	31.022,4	7.755,6
219.603,6	152.502,5	97.601,6	54.900,9	24.400,4
530.782,7	389.962,8	270.807,5	173.316,8	97.490,7
688.806,4	527.367,4	387.453,6	269.065,0	172.201,6
321.969,2	420.531,2	321.969,2	236.548,8	164.270,0
143.272,8	195.010,2	254.707,2	195.010,2	143.272,8
119.447,5	172.004,4	234.117,1	305.785,6	234.117,1
40.524,8	63.320,0	91.180,8	124.107,2	162.099,2
2.465.126,9	2.215.303,8	1.925.124,5	1.677.501,0	1.446.894,1

Segue: Calcolo delle A_l, cioè delle somme dei quadrati degli scost

	W	WSW	SW
E	155.411,2	118.986,7	87.418,8
ENE	139.875,4	182.694,4	139.875,4
NE	170.514,0	232.088,5	303.136,0
NNE	87.780,0	126.403,2	172.048,8
N	46.374,4	72.460,0	104.342,4
NNW	30.656,7	54.500,8	85.157,5
NW	20.927,6	47.087,1	83.710,4
WNW	6.575,6	26.302,4	59.180,4
W	..	7.755,6	31.022,4
WSW	6.100,1	..	6.100,1
SW	43.329,2	10.832,3	..
SSW	96.863,4	43.050,4	10.762.6
S	105.132,8	59.137,2	26.283.2
SSE	99.495,0	63.676,8	35.818,2
SE	172.004,4	119.447,5	76.446,4
ESE	124.107,2	91.180,8	63.320,0
$A_l =$ . . .	1.305.147,0	**1.255.603,7**	1.284.622,6

MENTI RISPETTO ALLE DIREZIONI – (*De Bilt*, Velocità del vento, 1925-1926).

SSW	*S*	*SSE*	*SE*	*ESE*
60.707,5	38.852,8	21.854,7	9.713,2	2.428,3
102.765,6	71.365,0	45.673,6	25.691,4	11.418,4
232.088,5	170.514,0	118.412,5	75.784,0	42.628,5
224.716,8	172.048,8	126.403,2	87.780,0	56.179,2
142.021,6	185.497.6	142.021,6	104.342,4	72.460,0
122.626,8	166.908,7	218.003,2	166,908,7	122.626,8
130.797,5	188.348,4	256.363,1	334.841,6	256.363,1
105.209,6	164.390,0	236.721,6	322.204,4	420.838,4
69.800,4	124.089.6	193.890,0	279.201,6	380.024,4
24.400,4	54.900,9	97.601,6	152.502,5	219.603,6
10.832,3	43.329,2	97.490,7	173.316,8	270.807,5
..	10.762,6	43.050,4	96.863,4	172.201,6
6.570,8	..	6.570,8	26.283,2	59.137,2
15.919,2	3.979,8	..	3.979,8	15.919,2
43.001,1	19.111,6	4.777,9	..	4.777,9
40.524,8	22.795,2	10.131,2	2.532,8	..
1.331.982,9	1.436,894,2	1.618,966,1	1.861.945,8	2.107.414,1

CALCOLO DELLE B_c, CIOÈ DELLE SOMME DEI QUADRATI DEGLI

	GENNAIO	FEBBRAIO	MARZO	APRILE	MAGGIO
Gennaio	..	8.117,9	32.471,6	73.061,1	129.886,4
Febbraio	7.861,5	..	7.861,5	31.446,0	70.753,5
Marzo	32.670,4	8.167,6	..	8.167,6	32.670,4
Aprile	59.621,4	26.498,4	6.624,6	..	6.624,6
Maggio	118.249,6	66.515,4	29.562,4	7.390,6	..
Giugno	147.202,5	94.209,6	52.992,9	23.552,4	5.888,1
Luglio	234.637,2	162.942,5	104.283,2	58.659,3	26.070,8
Agosto	142.255,0	204.847,2	142.255,0	91.043,2	51.211,8
Settembre	99.464,0	155.412,5	223.794,0	155.412,5	99.464,0
Ottobre	62.299,8	110.755,2	173.055,0	249.199,2	173.055,0
Novembre	29.178,0	65.650,5	116.712,0	182.362,5	262.602,0
Dicembre	8.263,3	33.053,2	74.369,7	132.212,8	206.582,5
B_c = . . .	941.702,7	936.170,0	963.981,9	1.012.507,2	1.064.809,1

…COSTAMENTI RISPETTO AI MESI — (*De Bilt*, Velocità del vento, 1925-1926).

Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre
202.947,5	292.244,4	20.947,5	129.886,4	73.061,1	32.471,6	8.117,9
125.784,0	196.537,5	283.014,0	196.537,5	125.784,0	70.753,5	31.446,0
73 508,4	130.681,6	204.190,0	294.033,6	204.190,0	130.681,6	73.508,4
26.498,4	59.621,4	105.993,6	165.615,0	238.485,6	165.615,0	105.993,6
7.390,6	29.562,4	66.515,4	118.249,6	184.765,0	266.061,6	184.765,0
..	5.888,1	23.552,4	52.992,9	94.209,6	147.202,5	211.971,6
6.517,7	..	6.517,7	26.070,8	58.659,3	104.283,2	162.942,5
22.760,8	5.690,2	..	5.690,2	22.760,8	51.211,8	91.043,2
55.948,5	24.866,0	6.216,5	..	6.216,5	24.866,0	55.948,5
110.755,2	62.299,8	27.688,8	6.922,2	..	6.922,2	27.688,8
182.362,5	116.712,0	65.650,5	29.178,0	7.294,5	..	7.294,5
297.478,8	206.582,5	132.212,8	74.369,7	33.053,2	8.263,3	..
…11.952,4	1.130.685,9	942.499,2	1.099.545,9	1.048.479,6	1.008.332,2	960.720,0

Scostamenti delle diverse coppie

	$d^2_{c_j}$ / $d^2_{l_i}$	Gennaio	Febbraio	Marzo	Aprile
		0	1	4	9
E	0	0	1	2	3
ENE	1	1	1.41	2.24	3.16
NE	4	2	2.24	2.83	3.61
NNE	9	3	3.16	3.61	4.2
N	16	4	4.12	4.47	5
NNW	25	5	5.10	5.39	5.83
NW	36	6	6.08	6.32	6.71
WNW	49	7	7.07	7.28	7.6
W	64	8	8.06	8.25	8.5
WSW	49	7	7.07	7.28	7.6
SW	36	6	6.08	6.32	6.7
SSW	25	5	5.10	5.39	5.8
S	16	4	4.12	4.47	5
SSE	9	3	3.16	3.61	4.2
SE	4	2	2.24	2.83	3.6
ESE	1	1	1.41	2.24	3.1

I MODALITÀ DELLA COPPIA *E-Genn.*

Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre
16	25	36	25	16	9	4	1
4	5	6	5	4	3	2	1
4.12	5.10	6.08	5.10	4.12	3.16	2.24	1.41
4.47	5.39	6.32	5.39	4.47	3.61	2.83	2.24
5	5.83	6.71	5.83	5	4.24	3.61	3.16
5.66	6.40	7.21	6.40	5.66	5	4.47	4.12
6.40	7.07	7.81	7.07	6.40	5.83	5.39	5.10
7.21	7.81	8.49	7.81	7.21	6.71	6.32	6.08
8.06	8.60	9.22	8.60	8.06	7.62	7.28	7.07
8.94	9.43	10	9.43	8.94	8.54	8.25	8.06
8.06	8.60	9.22	8.60	8.06	7.62	7.28	7.07
7.21	7.81	8.49	7.81	7.21	6.71	6.32	6.08
6.40	7.07	7.81	7.07	6.40	5.83	5.39	5.10
5.66	6.40	7.21	6.40	5.66	5	4.47	4.12
5	5.83	6.71	5.83	5	4.24	3.61	3.16
4.47	5.39	6.32	5.39	4.47	3.61	2.83	2.24
4.12	5.10	6.08	5.10	4.12	3.16	2.24	1.41

La tabella così formata potrà impiegarsi a determinare la somma degli scostamenti assoluti della data serie rispetto a ciascuna coppia di modalità. Così, volendo il valore di tale somma rispetto alla coppia *W S W — Febb* si applichi la tabella stessa, scritta preferibilmente su carta trasparente, sul prospetto iniziale delle velocità del vento, in modo che la sua casella superiore a sinistra, contenente lo scostamento 0, si sovrapponga alla coppia *W S W — Febbr*; ciascun valore inscritto nella tabella costituirà lo scostamento da *W S W — Febb* di quella coppia di modalità rappresentata dalla casella sottostante del prospetto iniziale (o da quella che risulterebbe sottostante per effetto di una permutazione circolare delle linee e delle colonne del medesimo prospetto).

La somma dei prodotti di tali scostamenti per i rispettivi valori di T_{ij} (nel nostro caso: velocità del vento) sarà il valore di Θ cercato in *W S W — Febb*.

In tal modo si trova $\Theta\,(W\,S\,W - Febb) = 390.446{,}45$, e pertanto:

$$ {}^{1}\eta_{A} = \frac{390446{,}45}{84.954{,}7} = 4{,}60 $$

Matrimoni per mesi e per giorni della settimana celebrati in Roma nel triennio 1925-27 (1)

63. — Un altro esempio di serie statistica dipendente da due caratteri ciclici ci è offerto dal seguente prospetto in cui, per il triennio 1925-27, risultano enumerati i matrimoni a seconda del mese e a seconda del giorno della settimana in cui vennero celebrati.

(1) Anche questi dati vennero cortesemente forniti dal Prof. L. Maroi.

MATRIMONI PER MESI E PER GIORNI DELLA SETTIMANA

(*Roma*, 1925-27)

GIORNO / MESE	LUNEDÌ	MARTEDÌ	MERCOLEDÌ	GIOVEDÌ	VENERDÌ	SABATO	DOMENICA	$L_i =$
Gennaio	189	6	119	196	8	223	406	1.147
Febbraio	361	16	136	230	11	261	507	1.522
Marzo	163	9	87	171	10	105	279	824
Aprile	475	3	257	446	7	329	375	1.892
Maggio	153	6	87	146	2	175	340	909
Giugno	373	14	201	279	7	281	462	1.617
Luglio	234	8	168	242	10	239	387	1.288
Agosto	281	1	136	208	4	226	354	1.210
Settembre	193	9	168	258	1	245	361	1.235
Ottobre	515	9	191	454	4	582	812	2.567
Novembre	148	5	73	157	5	190	300	878
Dicembre	246	12	169	295	4	187	390	1.303
$C_j =$	3.331	98	1.792	3.082	73	3.043	4.973	16.392

a) Con lo stesso procedimento indicato al numero 62 si è calcolato il momento secondo della data serie in corrispondenza a ciascuna coppia di modalità e si è ottenuto il prospetto seguente:

		Lunedì	Martedì	Mercoledì	Giovedì	Venerdì	Sabato	Domenica
	A_l \ B_c	52.806	65.387	88.940	89,165	64.046	47.708	50.924
Gennaio.	201.111	253.917	266.498	290.051	290.276	265.157	248.819	252.035
Febbraio	204.413	257.219	269.800	293.353	293.578	268.459	252.121	255.337
Marzo	211.459	264.265	276.846	300.399	300.624	275.503	259.167	262.383
Aprile	221.649	274.455	287.036	310.589	310.814	285.695	269.357	272.573
Maggio	203.015	255.821	268.402	291.955	292.180	267.061	250.723	253.939
Giugno	196.093	248.899	261.480	285.033	285.258	260.139	243.801	247.017
Luglio	190.693	243.499	256.080	279.633	279.858	254.739	238.401	241.617
Agosto	190.529	243.335	255.916	279.469	279.694	254.575	238.237	241.453
Settembre	186.631	239.437	252.018	275.571	275.796	250.677	**234.339**	237.555
Ottobre	195.741	248.547	261.128	284.681	284.906	259.787	243.449	246.665
Novembre	192.237	245.043	257.624	281.177	281.402	256.283	239.945	243.161
Dicembre	199,681	252.487	265.068	288.621	288.846	263.727	247.389	250.605

Si rileva da esso che il momento secondo risulta minimo per la coppia di modalità *Settembre-Sabato*, la quale deve quindi essere riguardata come media aritmetica della data serie (quando ci si limiti a considerare, come nei due esempi precedenti, le sole modalità effettive dei due caratteri ciclici da cui dipende la serie).

Si deduce subito che:

$$^{2}\eta_{A} = \left(\frac{234.339}{16.392}\right)^{\frac{1}{2}} = 3{,}78$$

Se poi si desidera conoscere anche lo scostamento semplice medio dalla media, basterà costruire il prospetto degli scostamenti assoluti delle varie coppie di modalità da una di esse, p. es. dalla *Genn* — *L*, e utilizzare tale prospetto, come già fu indicato al numero precedente, per il calcolo della somma Θ degli scostamenti assoluti rispetto alla coppia di modalità costituente la media aritmetica. Si trova allora Θ (*Sett* — *Sab*) = 55.943,57 e quindi

$$^{1}\eta_{A} = \frac{55.943{,}57}{16.392} = 5{,}84$$

b) Passiamo a determinare la mediana della serie data, cioè la coppia di modalità rispetto a cui risulta minima la somma (ponderata) Θ degli scostamenti assoluti delle altre coppie di modalità. Nello stesso modo che già servì a determinare la somma degli scostamenti assoluti rispetto alla media aritmetica, si trova che il valore di Θ per ciascuna coppia di modalità risulta quale è indicato dal seguente prospetto:

Somma Θ degli scostamenti assoluti da ciascuna coppia di modalità alle rimanenti

Mese \ Giorno	Lunedì	Martedì	Mercoledì	Giovedì	Venerdì	Sabato	Domenica
Gennaio.	59.430,13	61,524,82	64.468,14	64.352,64	61.565,74	58.469,30	58.569,81
Febbraio	59.746,23	61.457,86	64.843,09	64.782,75	62.041,13	58.540,91	58.988,60
Marzo	60.444,31	62.384,49	65.140,84	65.111,20	62.642,41	59.649,12	59.749,37
Aprile	61.004,04	62.867,99	65.879,63	65.808,24	63.308,12	60.151,27	60.488,32
Maggio	59.503,65	61.365,75	64.155,95	64.139,76	61.680,70	58.619,74	58.816,53
Giugno	58.632,40	60.387,15	63.632,37	63.583,75	60.885,52	57.461,96	57.899,90
Luglio	58.018,38	60.034,06	62.967,29	62.875,76	60.183,60	57.035,33	57,250,87
Agosto	57.749,15	59.477,46	62.917,75	62.761,31	59.873,15	56.251,67	56.893,03
Settembre	57.170,02	59.377,48	62.378,45	62.142,41	59.265,25	**55.943,57**	56.366,56
Ottobre	57.647,02	60.003,65	63.251,40	62.894,05	59.918,48	56.232,00	56.511.01
Novembre	57.979,82	60.099,31	63.229,32	62.991,19	60.032,44	56.694,49	56.998,55
Dicembre	59.015,67	60.813,75	64.133,20	63.938,89	61.097,74	57.544,10	58.135,12

Il valore di Θ risulta minimo ed uguale a 55943,57 per la coppia di modalità *Settembre–Sabato*, che costituisce dunque la mediana della serie. Lo scostamento semplice medio dalla mediana è quindi:

$$^{1}\eta_{M} = \frac{55.943,57}{16,392} = 5,84.$$

mentre lo scostamento quadratico medio si avrà osservando, dal prospetto prima costruito, che il momento secondo per la stessa coppia di modalità è 234,339, e che perciò

$$^{2}\eta_{M} = \left(\frac{234.339}{16.392}\right)^{\frac{1}{2}} = 3,78$$

Si noti che, in questo esempio, la coppia di modalità *Sett–Sab* costituisce ad un tempo la media e la mediana della serie.

c) Si voglia infine calcolare la differenza media della stessa serie, con ripetizione Δ_R e senza ripetizione Δ.

È chiaro che se in una serie dipendente da un carattere ciclico X ad s modalità e da uno Y a t modalità (equispaziate) è T_{ij} il valore della serie corrispondente alla modalità i di X ed j di Y, mentre $\Theta\,(i, j)$ è la somma ponderata degli scostamenti assoluti della coppia di modalità (i, j) dalle rimanenti, sarà

$$\Delta_R = \frac{1}{n^2} \sum_i \sum_j T_{ij} \Theta(i,j) \qquad \Delta = \frac{1}{n\,(n-1)} \sum_i \sum_j T_{ij} \Theta(i,j)$$

Perciò, nell'esempio considerato, in cui $n = 16.392$, T_{ij} è il numero dei matrimoni celebrati nel mese i e nel giorno j, e in cui il prospetto ultimamente calcolato fornisce il valore di Θ per ogni coppia di modalità, sarà

$$\Delta_R = \frac{981.166.291,73}{16.392^2} = 3,65 \qquad \Delta = \frac{981.166.291,73}{16.392 \times 16.391} = 3,65$$

d) L'esame del prospetto iniziale ci mostra che, corrispondentemente al massimo assoluto 812 della distribuzione, la moda o norma della serie è costituita dalla coppia di modalità *Ottobre-Domenica*, in cui risulta massimo il numero di matrimoni. E poichè gli altri due prospetti formati ci indicano che per la stessa coppia la somma Θ degli scostamenti assoluti è 56511,01, ed il momento secondo è 246.665, così risulta senz'altro che

$$^{1}\eta_{N} = \frac{56511,01}{16.392} = 5,87 \qquad ^{2}\eta_{N} = \left(\frac{246.665}{16.362}\right)^{\frac{1}{2}} = 3,08.$$

RIASSUNTO

I procedimenti di calcolo, immediatamente applicabili ai caratteri quantitativi ed alle serie statistiche che ne dipendono, consentirono agli statistici di dare alle teorie inerenti a tali serie un più ampio e più rapido sviluppo di quanto, fino a pochi anni addietro, non fosse stato possibile per le serie che dipendono da un carattere qualitativo. Per queste ultime, che a prima vista si sottraevano a quel potente mezzo di indagine e di sistemazione logica che è l'analisi matematica, la teoria rimase per lungo tempo allo stato di abbozzo.

Il GALTON, e dopo di lui il PEARSON, lo YULE, il BENINI riconobbero la possibilità di applicare anche ad esse, con opportuni accorgimenti, lo strumento analitico. Ma progressi decisivi in questo senso furono realizzati soltanto recentemente, specie ad opera di statistici Italiani (GINI, PIETRA): *a*) per effetto di una razionale distinzione dei caratteri qualitativi (in caratteri rettilinei, ciclici, sconnessi) (1); *b*) per avere riconosciuto che alcuni, almeno, dei procedimenti di calcolo applicati usualmente alle serie dipendenti da un carattere quantitativo, non esigono veramente la misurabilità di tale carattere, ma soltanto quella delle « diversità » fra le sue modalità (1). E poichè è possibile, regolandosi opportunamente, « misurare » anche le « diversità » fra le modalità di un carattere qualitativo, così ecco prospettarsi senz'altro la possibilità di indagare per via matematica anche le serie dipendenti da caratteri qualitativi.

* * *

Un punto sul quale non era stato fin qui posto attenzione, se si eccettua una chiara, per quanto succinta indicazione del senso in cui una eventuale indagine poteva essere fatta (2), era quello di vedere se e come potessero estendersi alle serie dipendenti da un carattere qualitativo (mutabile) i concetti delle diverse medie solitamente definite per le serie dipendenti da un carattere quantitativo (variabile) e i concetti dei diversi indici di mutabilità connessi all'esistenza di tali medie.

(1) C. GINI. *Varabiilità e Mutabilità*, già citato.
(2) C. GINI. *The Contributions of Italy to modern Statistical Methods*, già citato.

Sorvolando sul caso più semplice delle serie rettilinee, le quali si investigano in modo del tutto analogo alle serie dipendenti da un carattere quantitativo o seriazioni, e consentono quindi di accogliere per esse, con poche varianti, le consuete definizioni delle medie e degli indici di mutabilità che ne dipendono; sorvolando altresì su quella particolare media che è la moda o norma, della quale si può in ogni caso conservare la consueta definizione — rimaneva sempre aperta la questione di definire la media aritmetica e la mediana nei riguardi delle serie cicliche e delle sconnesse. Alla risoluzione di tale questione, è appunto rivolto, principalmente, il presente lavoro, il quale utilizza sistematicamente a tale scopo quel « principio di conservazione delle leggi formali » che è di così larga e feconda applicazione nell'analisi matematica.

È, difatti, evidente che le consuete definizioni di media aritmetica e di mediana, le quali presuppongono l'esistenza di due modalità estreme fra cui queste medie debbono essere comprese, non sono applicabili nè alle serie cicliche, nè alle sconnesse; ma se alle accennate definizioni altre se ne sostituiscono, costituite dal verificarsi delle proprietà che caratterizzano la media aritmetica o la mediana quando le definizioni tradizionali non cadano in difetto, allora non soltanto risulta la possibilità di stabilire i concetti di media aritmetica e di mediana per le dette serie, ma altresì quella di ottenere diverse estensioni a seconda che ci si basi sull'una o sull'altra proprietà formale, in surrogazione della definizione consueta; salvo poi a vedere quali estensioni risultino fra loro coincidenti. È da notarsi la circostanza che, in relazione a ciascuna definizione, risultano generalmente determinate non una sola, ma diverse modalità medie.

Il principio di conservazione delle leggi formali si può anche applicare in un altro senso, e cioè cercando se la media aritmetica di più numeri, interpretata meccanicamente come ascissa di quel punto di una retta a cui si applica la risultante di più forze parallele passanti per i punti della retta stessa che hanno come ascisse quei numeri, possa ricevere un'analoga interpretazione nei riguardi delle serie cicliche e delle sconnesse; ed effettivamente si trova che ciò è possibile per quelle, ma non per queste ultime serie.

È poi facile intendere come le estensioni dei concetti di media aritmetica e di mediana per le serie cicliche e per le sconnesse creino la susseguente possibilità di definire per queste serie degli indici di mutabilità dipendenti da tali medie. Naturalmente la possibile molteplicità di modalità medie in relazione a ciascuna definizione,

porta ad una pari molteplicità di indici di mutabilità, nel senso, per es., che se una serie ciclica ammette diverse medie aritmetiche, ammetterà anche uno scostamento semplice medio ed uno scostamento quadratico medio, corrispondentemente a ciascuna di tali medie.

Infine un'altra estensione si è naturalmente presentata. Fra le serie statistiche dipendenti da più caratteri quantitativi o qualitativi (serie a più dimensioni), ci siamo soffermati a considerare particolarmente quelle che sono funzioni di due soli caratteri. Per tali serie (di cui varie specie possono esistere, conformemente alla distinzione dei caratteri in quantitativi e qualitativi, ed alla suddivisione di questi ultimi in rettilinei, ciclici, sconnessi) ci è stato possibile stabilire dei criterî in base ai quali misurare la diversità fra due coppie di modalità di quei due caratteri da cui esse dipendono; dopo di che è stata facilmente eseguita, sempre con applicazione del principio di conservazione delle leggi formali, l'estensione alle serie stesse dei diversi concetti di media e quella dei diversi indici di variabilità o di mutabilità, sia indipendenti che dipendenti da tali medie. Da ciò abbiamo, in particolare, tratto occasione a rettificare il concetto di « centro mediano » della popolazione, così come tale centro viene da alcuni definito e determinato.

SUMMARY

The methods of calculation directly applicable to quantitative characters and to the statistical series dependent thereon, have enabled statisticians to develop the theories relating to such cases more fully and rapidly than had previously been possible in the case of series depending on a qualitative character. At first sight it did not seem possible to apply to the latter the powerful instrument of enquiry and of logical systematization afforded by mathematical analysis, and the theory long remained a mere outline as far as they were concerned.

GALTON and, after him, PEARSON, YULE and BENINI, recognised the possibility of applying the analytical method, with due precautions, to these series also.

But marked progress in that direction has only recently been made, chiefly as a result of the work of Italian statisticians (GINI, PIETRA): *a*) by drawing a rational distinction between the qualitative characters, (rectilineal, cyclical, unconnected); *b*) by recognising that at least some of the methods of calculation usually applied to series depending on a quantitative character, do not require that the character itself but only that the « diversities » between its modalities be measurable (1). And as it is possible by taking the requisite steps, to measure the « diversities » between the modalities of a qualitative character, it is found possible to apply the methods of mathematical analysis also to series depending on qualitative characters.

* * *

A point which had hitherto escaped attention – except for a clear but brief indication of the direction in which an eventual enquiry could be made – (2), was that of seeing if and how it would be possible to extend to series depending on a qualitative character (mutable) the concepts of several means usually determined for series depending on a quantitative character (variable) and the concepts of several indices of mutability arising from the existence of such means.

(1) C. GINI. *Variabilità e mutabilità*, op. cit.

(2) C. GINI. *The Contributions of Italy to Modern Statistical Methods*, op. cit.

Leaving on one side the simpler case of rectilineal series, which are studied on lines similar to those adopted for series depending on a quantitative character (or seriations), and to which we can therefore apply with few variations the usual definitions of means and of the indices of mutability depending thereon; leaving likewise on one side that special mean which is the mode, the usual definition of which holds good anyhow – the question of defining the arithmetic mean and the median in relation to cyclical and unconnected series still remained open.

It is to the solution of that question that the present study is mainly devoted, and for this purpose it systematically employs that « principle of the maintenance of formal laws » so widely and usefully applied in mathematical analysis.

It is indeed evident that the usual definitions of the arithmetic mean and of the median, which presuppose the existence of two extreme modalities, between which these means must be contained, cannot be applicable to cyclical or to unconnected series; but if the aforesaid definitions are replaced by others resulting from the ascertainment of the properties pertaining to the arithmetic mean or to the median when the usual definitions hold good, then not only does it become possible to determine for these series the concepts of the arithmetic mean and of the median, but it also becomes possible to secure diverse extensions, taking as a basis one or other formal propriety, in lieu of the usual definition subsequently examining which of the extensions coincide among themselves.

It should be noted thad it is usually found with regard to each definition that not one, but several average modalities have been determined.

The principle of maintaining of formal laws, is also susceptible of another application, i. e. by ascertaining whether the arithmetic mean of several numbers, mechanically interpreted as the abscissa of that point of a straight line, to which is applied the sum of several parallel forces passing through the points of the said straight line having those numbers as abscissas, is susceptible of a similar interpretation as regards the cyclical and the unconnected series; such enquiry shows indeed that this is possible for the former but not for the latter series.

It will be readily understood that the extension of the concept of the arithmetic mean and of the median, to cyclical and to unconnected series, makes it possible to determine for them the

indices of mutability, depending on the said means. Of course the possible plurality of average modalities in relation to each definition leads to a like plurality of the indices of mutability,viz, if a cyclical series allows of several arithmetic means it will also allow of a simple mean deviation and a quadratic mean deviation, corresponding to each of those means.

Lastly, another extension of this concept, has naturally arisen.

Among the statistical series, depending on several qualitative or quantitative characters (series of several dimensions) we have devoted special attention to those which are functions of two characters only.

For such series (of which there may be several kinds in conformity with their quantitative or qualitative characters and the subdivision of the latter into rectilineal, cyclical and unconnected) we have been able to lay down rules by wich to measure the diversities between two pairs of modalities of those two characters on which they depend: subsequently, and by applying the principle of the maintenance of formal laws, the diverse concepts of the mean and the concept of the diverse indices of variability or mutability, whether dependent or not on the said means, can be readily extended to the same series.

This has afforded us the opportunity of rectifying the concept of the « median point » of population as defined and determined by some.

G. E. DARMOIS

ANALYSE ET COMPARAISON DES SERIES STATISTIQUES QUI SE DEVELOPPENT DANS LE TEMPS (THE TIME CORRELATION PROBLEM)

1. Le problème dont nous allons nous occuper est particulièrement important en économie politique, mais il se présente dans beaucoup d'autres domaines de l'activité scientifique. Etudié depuis très longtemps, il n'a été approfondi que depuis peu.

Il n'y a guère qu'une trentaine d'années que la question a été posée avec une certaine précision. Elle a fait depuis des progrès considérables, tant pour la netteté des idées fondamentales que par l'efficacité des méthodes dirigées vers sa solution.

En 1921, G. Udny Yule fixait ainsi le but (1):

« Elucider les relations qui existent entre deux quantités qui varient dans le temps ».

Il peut arriver d'ailleurs qu'on ait à élucider aussi les relations entre les divers termes d'une série statistique observée dans le temps.

En somme, nous pouvons dire qu'il faut:

Analyser la structure et les liaisons internes d'une telle série A.

Etant données deux séries A, B, analyser ces deux séries et les liaisons qu'elles peuvent présenter entre elles.

Il s'agit là, on le voit, d'un vaste problème, posé en termes généraux et par là même un peu vagues. On ne doit pas être surpris des efforts qu'il a fallu dépenser pour préciser à la fois le problème et sa solution. Nous parlerons tout à l'heure des travaux qui nous paraissent les plus importants et qui ont fait faire à la question des progrès substantiels.

Mais on comprendra mieux, je pense, la direction de ces efforts et la signification de ces divers travaux si l'on replace d'abord le

(1) J. R. S. S. Vol. LXXXIV, pp. 497–537. *On the time correlation problem ...*

problème dans le grand ensemble que constitue la recherche des lois scientifiques.

Nous y gagnerons en outre, du moins je l'espère, de voir se dissiper les légers brouillards qui obscurcissent encore certains points de ce domaine, et disparaître l'allure paradoxale de quelques résultats.

Il nous suffira pour cela de revenir vers les principes et les définitions précises.

2. *L'observation scientifique – Description – Explication.* — Le but du travail scientifique est de mettre de l'ordre dans les résultats de l'observation. La foi du chercheur, peut on dire, est que cet ordre existe et qu'il peut être trouvé. Parti de cette conviction, il poursuit d'abord la formation d'un ordre descriptif, qui peut n'être que de pure commodité, mais qui lui permette d'embrasser aisément des ensembles. Ensuite, il cherche à discerner des rapports constants, certaines permanences.

La base de la science est que ces permanences existent. On cherche alors plus loin, car (1): « dès l'instant que des rapports cons-
« tants sont observés, l'esprit se refuse à croire que les permanences
« qui les caractérisent sont purement empiriques, il veut y voir la
« manifestation de lois scientifiques susceptibles d'être expliquées
« par un enchaînement logique des phénomènes...

On vise alors à l'ordre explicatif ou logique, à une théorie du phénomène.

3. *Permanences fonctionnelles – Lois fonctionnelles.* — Les grands succès de la science moderne, à partir du XVII^e^ siècle, ont été obtenus dans l'observation et la théorie des permanences fonctionnelles. Dans le cas simple de deux grandeurs mesurables, l'observation révélera que si l'une est fixée, l'autre l'est aussi. La pression et le volume d'une masse gazeuse (observée à température constante) présentent une telle liaison.

Nous dirons qu'il y a lien fonctionnel ou rigide entre les grandeurs.

La portée d'un projectile est liée rigidement à la vitesse initiale et à l'angle de tir. D'innombrables exemples analogues sont fournis par l'observation. La mécanique rationnelle a fourni de ces permanences fonctionnelles une explication d'ensemble; par l'introduction

(1) DIVISIA, *Economique rationnelle.* Paris 1928, p. 5.

de concepts, point matériel, vitesse, accélération, masse, force... elle a permis de dérouler un ordre logique qui reproduit l'ordre observé, avec une approximation excellente.

On peut dire que le développement dans le temps d'un phénomène mécanique à peu près quelconque est connu avec une précision très grande, dès que les conditions initiales et les forces sont connues.

4. *Les permanences statistiques.* — Dès la fin du XVII^e^ siècle et le début du XVIII^e^, la Statistique tirait de ses descriptions un assez grand nombre de régularités frappantes. Des rapports apparaissaient comme à peu près constants, dès que les observations portaient sur un grand nombre d'individus.

Cette stabilité de certaines fréquences (1), nous pouvons l'appeler une permanence statistique simple.

On sait comment JACQUES BERNOULLI, désireux de trouver une explication logique de ces permanences, y parvint à l'aide de la notion abstraite de probabilité, et fournit aussi la première théorie de ces régularités si spéciales.

Il faut insister un peu sur l'ordre nouveau qui s'introduit alors, sur les motifs que nous avons d'être satisfaits des succès de la théorie des probabilités, dans l'explication d'une longue série de pile ou face par exemple.

C'est au fond par l'analyse du mécanisme du jeu que nous parvenons à la probabilité du coup isolé, c'est parce que nous avons assez bien compris toute les circonstances possibles que nous lui fixons une valeur numérique.

Cette analyse approfondie du coup isolé, la façon dont nous appliquons les principes des probabilités, tout cela constitue une véritable théorie du phénomène.

De même, la stabilité des fréquences d'apparition des boules blanches d'une urne à composition donnée résulte pour nous d'une analyse du mécanisme de l'extraction.

5. *Permanences statistiques à une variable.* — Si l'on observe une population relativement à un caractère capable de variation, on obtient des résultats plus généraux. Un groupe nombreux détaché de la population manifeste une certaine répartition du caractère

(1) J. M. KEYNES. *A treatise on probability*, p. 336.

envisagé. Cette répartition est sensiblement la même pour deux groupes différents, s'ils sont tous deux assez nombreux.

Ces régularités dont QUÉTELET mit en évidence un grand nombre, peuvent être appelées des permanences statistiques à une variable. Leur étude est un des objets importants de la statistique.

On peut dire aussi que les courbes de fréquence des groupes nombreux sont généralement stables.

La théorie des probabilités fournit encore, comme on le sait, le correspondant abstrait de ces régularités observées, à l'aide de la notion de variable aléatoire possédant une loi de probabilité.

On arrivera dans ces cas à une véritable explication, qui sera plus ou moins complète, s'il est possible d'assigner *a priori*, par un raisonnement fondé sur des hypothèses raisonnables, certaines propriétés de la loi de probabilité.

La théorie des erreurs, dans les cas où elle réussit, est une explication de cette nature de la répartition observée des erreurs expérimentales.

6. *Permanences statistiques à deux ou plusieurs variables.* — Enfin les travaux de GALTON et de son école ont fait connaître des types plus complexes, et bientôt généralisés, de régularités statistiques.

Si les individus appartenant à une population sont classés relativement à deux caractères, deux groupes très nombreux fournissent généralement, l'expérience le montre, deux répartitions très voisines.

Nous dirons qu'on observe dans ce cas des permanences statistiques à deux variables, ou que les surfaces de fréquence sont stables.

Il faut alors distinguer deux cas. Ou bien, en fixant la valeur de l'un des caractères à ses diverses valeurs, on obtient pour l'autre caractère des répartitions qui sont toutes les mêmes.

Ou bien cette répartition varie avec la valeur donnée au premier caractère. C'est ce deuxième cas, tout nouveau, que GALTON rencontra et dont il commença l'étude. En somme, pour une valeur donnée au caractère x, le caractère y possède une loi de distribution stable, fixée avec x, mais variant avec lui.

Nous avons là une combinaison des notions déjà rencontrées. Les permanences statistiques à une variable y sont en liaison fonctionnelle avec x. C'est ce qu'à la suite de GALTON, on appela des corrélations.

La théorie des probabilités fournit la correspondance abstraite en considérant deux variables aléatoires x et y, la loi de probabilité de y, quand x est connu, dépendant fonctionnellement de x.

La loi de probabilité de x considéré seul est d'ailleurs supposée connue. Cela revient, comme on sait, à se donner la loi de probabilité à deux variables x et y.

Une explication partielle, ou complète, des régularités observées sera obtenue dans cet ordre d'idées si des raisonnements a priori peuvent faciliter ou permettre le calcul de la loi de probabilité.

Une telle dépendance que, relevant un mot de JACQUES BERNOULLI, on peut qualifier de stochastique (1), présente une très grande importance. Il s'agit d'un véritable lien entre variables aléatoires, lien analogue au lien fonctionnel, mais plus souple en quelque sorte, restreignant seulement la liberté d'une variable aléatoire quand est fixée la valeur de l'autre.

La dépendance stochastique ou lien de probabilité est donc une notion abstraite, s'appliquant aux êtres mathématiques que sont les variables aléatoires.

7. — On emploie aussi très fréquemment le mot de corrélation, et l'on dit que deux variables aléatoires sont en corrélation, pour exprimer la même idée de lien de probabilité.

On est donc exposé à donner le même nom à deux choses assez différentes:

1) au fait expérimental que les courbes de fréquence du caractère y pour x donné, varient avec la valeur de x; y est en corrélation avec x;

2) au fait mathématique que la loi de probabilité de la variable aléatoire y, pour x donné, varie avec la valeur de x.

Le lien entre ces idées est le suivant. Si, dans la population envisagée on choisit, comme on le dit d'habitude, un individu au hasard, c'est-à-dire par une méthode indifférente aux caractères étudiés, on obtiendra pour l'individu extrait, sous les hypothèses faites, un couple de variables aléatoires. La loi de probabilité de

(1) TCHOUPROW. *Grundgbegriffe und Grundprobleme der Korrelationstheorie* p. 20. E. SLUTSKY. *Sur les fonctions éventuelles continues,... dans le sens stochastique* C. R. Ac. Sc., T. 187 1928 pp. 878-880

Pour les fondements de la notion de limite stochastique, voir F. P. CANTELLI « Rendic. del Circolo Mat. di Palermo », T. 111-1916 pp. 191-201.

ces variables est identique à la loi de fréquence qui se trouve réalisée en fait.

S'il y a corrélation au sens 1) il y a corrélation au sens 2).

D'autre parti si l'on extrait des groupes nombreux, chacun d'eux présentera au sens 1) une corrélation et les lois de ces corrélations manifesteront une permanence, elles ressembleront beaucoup à la correlation du groupe total. L'explication de cette permanence se trouve dans l'existence d'une corrélation au sens 2).

Mais la corrélation au sens 1) dans le groupe total ne pourrait être expliquée que un nouvel appel à une autre corrélation au sens 2).

Si, pour un moment, nous appelons corrélation de fait la constatation expérimentale 1), on peut dire qu'une corrélation de fait est la base d'une corrélation (au sens abstrait) qui explique la permanence des corrélations de fait des groupes.

Généralement on essaie de vérifier la permanence des corrélations de fait des divers groupes pour en induire la corrélation de fait du groupe total.

Il est rare qu'on atteigne l'explication de cette corrélation de fait du groupe total.

Cette digression trop longue nous permettra du moins de voir clairement qu'entre les deux sens du mot corrélation, existe la même différence qu'entre la fréquence et la probabilité.

8. *Les caractères généraux de certaines séries statistiques.* — Le premier effort à faire, en présence d'une série statistique, est d'en tirer une description raisonnable, qui soit comme l'ébauche assez grossière d'une théorie.

Le phénomène soumis à l'observation est sous la dépendance de vastes ensembles de causes, d'une analyse sans doute difficile, mais qui s'offrent d'un bloc à l'esprit.

S'il s'agit de récoltes par exemple, on pensera à la qualité et à l'état des semences, à l'état du sol et à ses qualités naturelles, à la technique de la culture, au temps enfin.

On a quelque prise sur les premiers ensembles. Si l'on améliore les semences, le sol, les méthodes, la récolte va croître. Mais ces efforts d'amélioration peuvent être secondés ou contrariés par le temps de sorte qu'il paraît dès l'abord raisonnable d'essayer de distinguer ce lent effort qui tend à accroître la récolte des aléas qui peuvent l'augmenter ou la diminuer.

D'où la décomposition de la grandeur observée en deux composantes; l'une est la tendance générale qui, dans le cas envisagé doit fournir une courbe montante, l'autre est une composante aléatoire dont les valeurs sont tantôt positives, tantôt négatives.

Représentée à part si l'on pouvait l'isoler, cette composante aléatoire donnerait un graphique oscillant de part et d'autre de l'axe des temps. Bien entendu, ces premières réflexions ne prétendent qu'à esquisser très grossièrement et qualitativement les réactions très complexes qui mènent des améliorations réalisées à l'augmentation de la récolte.

Si l'on pouvait justifier que la vitesse d'accroissement de la récolte fût à peu près constante, la courbe de tendance serait une droite. Si les améliorations d'une année retentissent sur la suivante, on peut avoir au début une montée plus rapide, d'allure parabolique. On voit qu'on est conduit à tenter une description, raisonnable sans doute, mais assez grosse, et reposant sur une analyse peu approfondie.

Pourtant ces reflexions préliminaires sont une aide précieuse.

Si par exemple, le temps s'est montré exceptionnellement favorable, on évitera de faire monter trop la courbe de tendance.

9. — Imaginons alors que, nous entourant de tous les renseignements sur le sujet étudié, nous ayons dégagé une courbe de tendance. Elle représentera pour nous ce qu'auraient donné les récoltes si les conditions météorologiques étaient restées à peu près les mêmes dans les années successives.

Il reste une composante aléatoire, dont la moyenne générale doit être sensiblement nulle sur une période assez longue, et qui représente les effets des bonnes et des mauvaises années.

Une idée assez naturelle est que ces bonnes ou mauvaises années, sans suivre une loi déterminée en fonctions du temps, possèdent peut-être une certaine loi de fréquence. C'est par une série de variables aléatoires convenablement choisies, relatives à la température, à la pluie, au vent... qu'on représentera la physionomie météorologique d'une année. La variation de la production, conséquence des fluctuations dans la physionomie de l'année, est elle-même une variable aléatoire. *Supposons qu'elle possède une certaine loi de fréquence.*

Tout ceci paraît en première approximation assez satisfaisant.

Maintenant, la succession des valeurs, supposées exactement observées, de cette variable aléatoire, doit pouvoir nous renseigner sur cette loi de fréquence; avec quelle précision?

Il est alors utile de se demander si ces valeurs successives sont indépendantes au sens des probabilités.

Quand on sait qu'une année a été bonne, cela change-t-il les chances d'avoir ensuite une bonne ou une mauvaise année?

Il paraît d'abord raisonnable de supposer qu'il y a indépendance.

Nous aurions donc à traiter une série de valeurs indépendantes d'une même variable aléatoire.

Bien entendu, cette vue qui paraît adaptée à un problème particulier, ne saurait être conservée pour d'autres sans examen préalable. Même ici, il ne faut l'accepter que comme une hypothèse qui facilite les calculs.

Car qui nous dit qu'une année très sèche et par là désastreuse, ne peut avoir pour effet de préparer le sol pour l'année suivante, de sorte qu'un temps favorable conduise à des accroissements exceptionnels. La variable aléatoire causale se présenterait en séries indépendantes, mais les effets seraient des variables liées entre elles.

Bien d'autres circonstances peuvent être imaginées qui conduiraient à des résultats analogues.

10. *Etude plus précise d'une série statistique.* — Supposons analysée d'après les vues précédentes une série statistique. On peut essayer d'aller plus loin, de connaître ou de comprendre mieux le phénomène étudié. La courbe de tendance représente pour nous quelque chose comme les amplitudes successives d'un phénomène se développant dans des conditions qui restent constantes ou varient suivant certaines lois régulières. Nous pouvons désirer trouver pour ces amplitudes successives, une loi de forme mathématique en fonction du temps.

Notre but peut-être: 1) de pure description et de commodité, si une formule simple représente avec fidélité les observations; 2) plus élevé, viser à une représentation objective, à une induction, à une théorie. Expliquons nous sur un exemple qui touche à la fois à la physique et à la statistique, les taches du soleil. On sait qu'on peut déduire de ces taches un véritable indice de activité solaire. Les observations poursuivies depuis 1749 donnent pour cet indice un graphique frappant, un peu irrégulier sans doute, mais suggérant immédiatement l'idée d'un véritable phénomène périodique. On peut donc espérer représenter la partie régulière à l'aide d'un certain nombre de termes sinusoïdaux.

Comme autre exemple, imaginons que le phénomène étudié soit une croissance continue d'un organisme plus on moins complexe. Une connaissance même assez grossière du mécanisme de cette croissance peut conduire à poser une équation différentielle, et à essayer pour les courbes de tendance des courbes intégrales de cette équation. Le cas le plus simple, où la croissance est proportionnelle à la taille déjà acquise, conduite à la loi exponentielle.

En résumé, qu'il s'agisse seulement de décrire commodément, ou qu'il s'agisse de vérifier ou de poursuivre une théorie, on essaiera de donner de la courbe de tendance une expression mathématique en fonction du temps, loi linéaire ou parabolique, exponentielle, somme de sinus...

La composante aléatoire sera bien entendu, représentée pour nous par la différence entre l'ordonnée observée et l'ordonnée de la courbe de tendance ainsi conçue (1).

11. *La comparaison des séries statistiques.* — Venons maintenant à deux séries statistiques qu'on a pu soumettre à l'analyse précédente. Imaginons par l'exemple que l'une soit relative aux récoltes d'une région nettement agricole, l'autre se rapportant à la consommation du vin dans cette région. On peut se demander s'il existe une relation, et de quelle nature, entre les récoltes et la consommation du vin. Mais on voit déjà que cette question est vague et demande à être précisée.

Il peut exister une relation entre les tendances, l'amélioration continue des récoltes entrainant une augmentation continue de la consommation du vin. Il peut exister aussi, et cette relation serait assez naturelle dans certains cas, un lien entre les fluctuations de ces quantités.

Mais que signifie même ce terme de relation entre deux phénomènes figurés par deux courbes statistiques, et comment peut on mettre un tel lien en évidence?

(1) Pour une conception différente du rôle de la variable aléatoire, le lecteur peut se reporter à un travail de G. Udny Yule: *On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers.* « Ph. Trans. A. », Vol. 226, pp. 267–298.

On trouvera dans ce travail le graphique relatif aux taches du soleil.

Voir aussi un travail de H. Hotelling: *Differential equations subject to error....* « J. of the Am. Statist. Assoc. », Sept. 1927 pp. 283-314.

Envisageons les deux courbes de tendances. Quels points de ces courbes mettrons nous en correspondance? Dans le cas concret envisagé il paraît d'abord naturel de calculer les consommations d'une récolte à l'autre, et, ceci fait, de rapporter la consommation obtenue à la récolte qui l'a précédée. Les graphiques étant ainsi mis en regard, on regarde s'ils se ressemblent. Si vraiment ils se ressemblaient beaucoup, ce serait un résultat encourageant, et renforcerait l'idée que la consommation dépend de la récolte, idée qui n'est pas entièrement *a priori*, mais qui nous est suggérée par l'importance des récoltes dans les ressources de la population et par une connaissance suffisante des habitudes et de la psychologie des habitants. On peut alors essayer d'aller plus loin, de donner une forme plus précise, mathématique s'il est possible à cette relation entrevue. Ici encore, on peut poursuivre deux buts assez distincts:

1) Décrire par une formule simple un ensemble de résultats.

Il serait évidemment très commode, si les quantités étaient sensiblement proportionnelles, de résumer dans une seule constante de proportionnalité, toutes les observations.

2) Viser à la vérification d'une idée plus profonde. On peut avoir des raisons de croire dans certains cas que deux grandeurs statistiques sont deux aspects différents d'une même activité, deux indices d'un même complexe et peuvent être liés rigidement, c'est-à-dire par une relation fonctionnelle.

Une telle vue du problème ne peut être prise qu'après une étude assez approfondie; si on l'a vérifiée, même d'une façon satisfaisante, il ne faut pas perdre de vue les limitations auxquelles reste soumise l'utilisation d'une telle loi empirique.

Un exemple fera mieux comprendre notre pensée. Si nous observons une masse gazeuse dont le volume varie assez lentement, nous pourrons remarquer que le produit du volume par la pression reste constant, ou si l'on veut que la densité est proportionnelle à la pression. La densité et la pression de la masse gazeuse sont deux indices, qui ne diffèrent qu'en apparence, de l'état de cette masse, et ces deux indices sont proportionnels. Mais nous savons aussi que la constante de notre formule n'est pas une constante absolue, c'est une moyenne statistique qui représente la température absolue. C'est parce que celle-ci a peu varié dans nos expériences que nous avons trouvé notre loi. Si maintenant, confiant dans la solidité de notre vérification, nous comprimions brusquement le gaz, nous serons surpris de ne pas trouver le volume attendu.

Dans les lois que peut établir la statistique, figureront des constantes empiriques. Même si la loi est exacte, la signification de ces constantes nous échappera généralement. Il faut donc être prudent dans l'application de ces lois.

Il existe, si l'on ose dire, certaines températures économiques qu'on risque de faire varier brusquement.

Sous ces réserves d'interprétation et d'application, on est ramené à un problème analogue à celui du § 10.

Trouver une expression mathématique de l'une des grandeurs en fonction de l'autre.

12. *Représentation des séries fonctionnelles.* — Considérons d'abord une série isolée à représenter en fonction du temps. Pour fixer les idées, supposons qu'il s'agisse d'une courbe continue. S'il s'agissait d'un graphique de points isolés, les intégrales qui vont s'introduire seraient remplacées par des sommes. Soit donc une fonction $x\ (t)$ donnée par l'observation. Le cas le plus simple est celui où la courbe est à peu près droite. On cherche une fonction linéaire du temps, généralement par la condition du minimum de l'intégrale: $\int [x - a - bt]^2 dt$ étendue à l'intervalle $t_o, t_o + T$ des observations. On peut toujours supposer qu'on a retranché de x sa valeur moyenne $\frac{1}{T}\int x\ dt$ et qu'on a introduit un facteur tel que $\frac{1}{T}\int x^2\, dt = 1$. La fonction à représenter sera dite normée dans l'intervalle.

Dans ces conditions, a doit être nul. Si nous représentons par θ la fonction λt, λ étant tel que θ soit normée, on est ramené au minimum de: $\frac{1}{T}\int (x - p\theta)^2\, dt$, p étant l'inconnue.

On trouve immédiatement que le minimum a lieu pour:

$$p = \frac{1}{T}\int x\ (t)\ \theta\ (t)\ dt$$

et que sa valeur est $1 - p^2$, ce qui établit l'inégalité $1 - p^2 \geqslant 0$

Ce calcul peut être fait avec une fonction normée θ *absolument quelconque et donne les mêmes résultats.*

On reconnait dans p ce qu'on a l'habitude d'appeler le coefficient de corrélation des fonctions x et θ. Cet usage nous paraît regrettable; le mot de corrélation a sans doute dans la langue cou-

rante une signification très étendue, mais il faut ici lui donner un sens précis. Nous croyons qu'il vaudrait mieux, compte tenu des remarques du § 7, le réserver pour la dépendance de probabilité de deux variables aléatoires.

Ici, au contraire, le coefficient p possède déjà un nom en calcul fonctionnel. On sait en effet que si $p = O$, les fonctions sont dites orthogonales. Le coefficient p est donc une sorte de cosinus de l'angle des deux fonctions dans l'intervalle considéré; si l'on remarque qu'il n'atteint les valeurs ± 1 que lorsque l'intégrale est identiquement nulle, c'est-à-dire quand les fonctions sont proportionnelles, on peut aussi le considérer comme un indice de linéarité. S'il est très voisin de $+1$ ou de -1, les fonctions sont à peu près proportionnelles. Voyons d'une façon plus précise la signification de la valeur numérique de $1 - p^2$, carré moyen de la différence $x - p\theta$.

Que signifie en général l'égalité

$$\frac{1}{T}\int Z^2 dt = \varepsilon^2 \quad \text{pour une fonction } Z \text{ telle que } \frac{1}{T}\int z\, dt = o\,.$$

Il existe une analogie évidente entre cette question et la suivante:

Une série de nombres a sa moyenne nulle, son écart quadratique moyen est ε. Que peut-on en conclure?

La condition de la moyenne nulle exprime d'abord que l'écart quadratique moyen, pris par rapport à l'axe adopté, est inférieur à celui qu'on obtiendrait avec tout axe parallèle au premier.

D'autre part, si m et M sont les bornes inférieure et supérieure de la fonction Z, si $F(\lambda)$ représente la fraction de l'intervalle $t_0\ t_0 + T$ où Z est compris entre m et λ, il est clair que l'on a:

$$\varepsilon^2 = \int_m^M \lambda^2\, d\,F(\lambda) \qquad 0 = \int_m^M \lambda\, d\,F\lambda \qquad 1 = \int_m^M d\,F(\lambda)$$

m et M peuvent être $-\infty$ *et* $+\infty$.

$F(\lambda)$, qui n'est autre que la fonction sommatoire de Lebesgue relative à la fonction $Z(t)$ est une fonction de fréquence totale dont la différentielle est la fonction de fréquence élémentaire, et nous sommes ramenés à la question suivante:

Que sait-on d'une loi de probabilité dont la moyenne est nulle et l'écart quadratique moyen donné?

Rappelons ici le théorème de TCHEBICHEF, en l'énonçant dans le cas qui nous intéresse.

La fraction de l'intervalle total où la fonction Z dépasse U fois l'écart ε est inférieure ou égale à $\frac{1}{U^2}$.

Il en résulte qie si ε est très petit, la fonction Z ne peut prendre des valeurs notables que dans un ensemble de valeurs de t dont la mesure est très petite, mais il faut pour cela que ε soit très petit, c'est à dire que p sont extrêmement voisin de 1.

Imaginons $p = 0{,}8$ $\qquad \varepsilon^2 = 0{,}36 \qquad \varepsilon = 0{,}6$

On voit qu'on ne tirera pas grand chose du théorème de TCHEBICHEF, car le double de ε n'est exclus que sur le $\frac{1}{4}$ de l'intervalle, et le double de ε est très important pour une fonction dont l'écart quadratique moyen est l'unité.

Aussi la seule connaissance de ε est de peu de conséquence si ε n'est pas très petit. Il faudrait avoir en plus quelque idée de la loi de fréquence de la difference $x\text{-}p\theta$.

Remarque. — On pourrait croire que le problème à résoudre est plus particulier, car nous avons déterminé la fonction $Z(t) = x\text{-}p\theta$ par une condition de moindres carrés. $Z(t)$ n'est pas une fonction arbitraire. On voit aisément que $Z(t)$ est en effet orthogonale à la fonction θ. Mais cette condition est la seule qui lui soit imposée, avec la condition d'avoir sa moyenne nulle. Elle doit donc simplement être orthogonale aux deux fonctions 1 et θ. On reconnaît aisément que rien n'est changé aux conclusions précédentes.

13. *Représentation d'une fonction par combinaison linéaire de fonctions orthogonales.* — Nous allons rappeler maintenant quelques propriétés classiques (1).

Considérons un système de fonctions orthogonales et normées

$$\varphi_0(t) \; \varphi_1(t) \ldots \qquad \varphi_k(t) \ldots$$

satisfaisant par conséquent aux conditions:

$$\int \varphi_i \, \varphi_k \, dt = 0 \qquad \frac{1}{T}\int \varphi_i^2 \, dt = 1$$

(1) Voir p. ex. P. LÉVY. *Leçons d'analyse fonctionnelle.* Chap. VII, pp. 116–125. On y trouvera également p. 27 la définition de la fonction sommatoire $F(\lambda)$ utilisée plus haut.

Cherchons à approcher $x\ (t)$ par la combinaison linéaire

$$y\ (t) = a_0\,\varphi_0 + a_1\,\varphi_1 \ldots + a_m\,\varphi_m$$

de manière que l'intégrale suivante soit un minimum:

$$\frac{1}{T}\int (x - y)^2\,dt$$

En dérivant par rapport à a_i, on obtient immédiatement:

$$a_i = \frac{1}{T}\int x\,\varphi_i\,dt$$

et la valeur du minimum est alors:

$$1 - a_0^2 - a_1^2 \ldots - a_m^2$$

Si p_m^2 est la somme $\sum\limits_{i=0}^{m} a_i^2$ et p^2 la somme de la série évidemment convergente qu'on obtient en faisant croître m indéfiniment:

$$p_m^2 \leqq 1 \qquad p^2 \leqq 1$$

La suite des fonctions φ_i est dite complète s'il n'existe aucune fonction orthogonale à toutes les fonctions φ_i. On démontre alors que $p^2 = 1$. La suite est incomplète dans le cas contraire.

On sait que la suite normée:

$$1 \qquad \sqrt{2}\cos\ 2\pi t \qquad \sqrt{2}\sin\ 2\pi t \ldots \qquad \sqrt{2}\cos 2n\,\pi t \qquad \sqrt{2}\sin 2n\,\pi t \ldots$$

est complète dans l'intervalle 0 1.

La suite des polynômes de Legendre est également complète dans l'intervalle $-1 + 1$.

On voit qu'on peut, avec une approximation aussi grande qu'on le désire, représenter une fonction continue absolument quelconque par une combinaison linéaire de fonctions sinusoïdales ou de polynomes, puisque si m croît indéfiniment p_m^2 tend vers 1.

On peut dire que p_m, p sont des indices de linéarité de la fontion $x\ (t)$ par rapport au système des fonctions φ_i.

On voit qu'en prenant n'importe quel système complet, et allant assez loin, on est assuré d'obtenir un indice de linéarité très voisin de l'unité.

Bien entendu, il faut tenir compte de la simplicité de la représentation, et préférer celle qui comporte deux termes à celle qui en comporte une vingtaine. Si par chance, on obtient une bonne repré-

sentation avec un terme unique, on retombe sur ce qu'on appelle d'habitude le coefficient de correlation des deux fonctions (1).

14. *La signification d'une représentation.* — Que peut maintenant signifier le fait qu'on a obtenu une représentation simple et précise. Reportons nous à un remarquable travail de G. UDNY YULE «J. R. S. S.» Vol. LXXXIX, pag. 1-64 « Why do we sometimes get non sense correlations between time series... » et prenons l'exemple cité par l'auteur comme type de « non sense correlation ».

Les deux grandeurs observées sont la mortalité et la proportion des mariages faits par l'Eglise d'Angleterre. La période observée va de 1866 à 1911. Si l'on calcule l'indice de linéarité p, on trouve 0,9512.

Cette valeur est assez voisine de l'unité. D'ailleurs le graphique est frappant. Qu'est ce à dire?

L'une des grandeurs pourrait être calculée à partir de l'autre par une relation du premier degré, valable pour la période considérée. Il ne serait donc pas très hasardeux de dire:

Si je connaissais la mortalité de l'an prochain, je pourrais vous dire à peu près quelle sera la proportion des mariages de la catégorie envisagée. Nous sommes un peu choqués de cette affirmation. Nous avons tort. Cela ne veut rien dire de plus que ceci:

Les deux phénomènes étudiés manifestent tous deux une décroissance régulière, sensiblement linéaire par rapport au temps. Tant que se maintiendra cette régularité empirique, les phénomènes eux-mêmes auront des grandeurs en relation linéaire.

Cette relation linéaire n'est pas très précise. Nous savons que même un indice 0,9512 ne doit pas faire illusion. Mais le coefficient serait il plus voisin de l'unité, nous n'aurions pas encore devant nous une absurdité que notre raison doive combattre, et une véritable difficulté à élucider.

Il n'est nullement absurde en effet de penser que l'ensemble des mesures prises pour abaisser la mortalité ait agi d'une façon régulière et développe ses effets dans le temps de manière sensiblement linéaire.

(1) Voir dans: R. COURANT und D. HILBERT *Methoden der mathematischen Physik*. I. pp. 42–43, des méthodes de mesure de l'indépendance linéaire d'un groupe de fonctions.

Il n'est pas absurde non plus de supposer, dans l'évolution d'une société, qu'un ensemble de causes agisse pour diminuer l'autre phénomène, et que cette décroissance dans le temps soit linéaire.

Il en résulte donc que les deux grandeurs mesurées vérifient une relation linéaire. On peut même ajouter que non seulement ces grandeurs considérées la même année, mais prises avec des décalages constants, vérifient encore une relation linéaire.

Ce qui serait tout à fait absurde, c'est de penser qu'une telle relation linéaire, assez grossièrement vérifiée, valable pour un petit intervalle de temps, soit une base suffisante pour l'induction d'une loi reliant objectivement les deux grandeurs, alors que la moindre reflexion, non seulement ne nous incline pas vers une telle loi, mais nous démontre qu'il ne peut y avoir de liaison. Car il suffirait de concevoir une découverte scientifique diminuant brusquement la mortalité, et particulièrement accessible à un milieu social qui se marie à l'Eglise d'Angleterre, pour faire varier les grandeurs en sens inverse.

Ainsi, une représentation n'a pas de sens en elle-même. Elle peut n'être qu'un moyen commode de rassembler des observations.

Elle n'a pas à elle seule le pouvoir de nous pousser vers une induction quand la raison nous en écarte.

Si nous voulons pouvoir raisonner sur un tel résultat, il nous faut partir d'une base, si petite soit elle, il faut ensuite que le résultat observé manifeste sa permanence.

C'est pourquoi j'estime, étant donné la force de suggestion du seul mot de corrélation, qu'il vaudrait mieux renoncer dans ces questions à l'emploi du qualificatif coefficient de corrélation. En l'appellant indice de linéarité, ou quelque chose d'analogue, nous le ramenons à un rôle plus modeste et formel.

Et, par exemple, ce ne sera plus à cet indice de linéarité qu'il faudra faire un reproche si, considérant deux grandeurs que nous savons liées rigidement par une loi sinusoïdale, nous cherchons à les relier linéairement. Si nous devons alors être surpris, c'est d'avoir eu une idée aussi singulière.

En résumé, pour faire vraiment oeuvre explicative, il faut réfléchir sur le phénomène étudié, nous entourer de tous les renseignements qui puissent le faire comprendre, essayer de deviner quel système de fonctions paraît apte à le représenter.

Notre but doit être toujours de vérifier des rapports logiques ou plausibles, et si nous pouvons demander à des instruments de calcul de contrôler nos idées, il ne faut pas exiger qu'ils nous les fournissent.

15. *La composante aléatoire – Relation de deux composantes aléatoires.* — Nous obtenons d'abord une série de nombres:

$$x_1 \quad x_2 \ldots x_n$$

qui sont pour nous les valeurs observées d'une même grandeur aléatoire, dont nous voulons connaître la loi de probabilité.

Il s'agit donc d'un problème classique. On déterminera des valeurs empiriques des différents moments de cette loi de probabilité. Ces valeurs présumées sont affectées d'erreurs qu'on peut estimer assez aisément si les valeurs x_i sont indépendantes.

S'il s'agit de comparer deux composantes aléatoires, on cherchera la loi de probabilité à deux variables qui peut les relier.

D'après de que nous avons vu, l'idée la plus simple est que les valeurs x_i sont indépendantes, que les y le sont aussi. Nous supposerons de plus qu'une valeur x_i est liée à une seule valeur y_i, nous donnerons le même indice aux grandeurs que nous supposons liées. Deux valeurs x_i et y_k à indices différents sont donc indépendantes.

Ceci posé, nous possédons n couples d'une loi de probabilité à deux variables. La détermination complète de la loi demanderait le calcul de tous les moments, c'est-à-dire des quantités:

$$M_{p/q} = E\,(x^p\,y^q)$$

E désignant l'espérance mathématique.

Il arrive souvent qu'on se contente, les valeurs moyennes étant supposées nulles, des trois moments du second ordre. On peut toujours supposer les écarts types de x et y ramenés à l'unité, il ne reste alors que le moment $M_{1/1}$ qui se confond avec le coefficient de corrélation. Il est clair qu'une telle détermination de la loi de probabilité est généralement très insuffisante. On sait en effet que si les régressions ne sont pas linéaires, on s'expose à des erreurs notables en se contentant du coefficient de corrélation.

K. Pearson a le premier indiqué un moyen assez simple d'éviter ces difficultés, c'est d'étudier à côté de la loi de probabilité de x et de la loi de probabilité de y, indiquées par leurs écarts types, la loi de probabilité de *y pour x fixé*. C'est de la considération de cette loi liée que se tire le rapport de corrélation.

Tous ces points étant rappelés, nous allons maintenant parler des nombreux travaux consacrés à la question qui nous occupe.

16. *Les travaux – L'apparition des idées fondamentales.* — La recherche des relations entre grandeurs statistiques semble avoir

procédé tout d'abord d'idées assez confuses. On ne perçoit pas très nettement la nécessité de séparer la tendance générale des oscillations qui lui sont superposées. C'est dans un travail de R. H. HOOKER « J. R. S. S. » Vol. LXIV, 1901, p. 485 qui se trouve pour la première fois énoncée nettement l'idée qu'il est souvent indiqué de chercher les relations entre ces oscillations, et qu'il est par conséquent nécessaire d'éliminer le mouvement général.

Cette élimination est obtenue en calculant la différence de la valeur observée et de la moyenne de 9 années ayant l'année centrale au point d'observation.

Dans le même volume, le même auteur (p. 574) propose d'employer pour cette élimination du mouvement général, une méthode de différences.

Miss F. E. CAVE « Proc. R. Soc. » A. 74, 1904, p. 403, développe des idées analogues, HOOKER précise les siennes d'une façon satisfaisante dans « J. R. S. S. » Vol. LXVIII, 1905, p. 696. L'idée est acquise de déterminer ou d'éliminer la tendance séculaire, et d'isoler les oscillations, en somme de déterminer séparément deux composantes, dont les significations sont tout à fait différentes, mais la deuxième paraît correspondre à une sorte de phénomène d'oscillation autour d'une position d'équilibre mobile.

L. MARCH, dans « J. S. S. », Paris, 1905, p. 255 et 306, parvient indépendamment à des résultats semblables à ceux de HOOKER.

Les méthodes de HOOKER sont utilisées en France par BUNLE (1911) (1), par LENOIR, dans sa thèse (1913) (2).

C'est en 1914 qu'apparaît une idée nouvelle, tendant à justifier l'emploi des différences successives pour trouver la relation entre les oscillations.

« Student » dans « Biometrika » X, 1914, p. 179, considère les mouvements séculaires comme des phénomènes réguliers de développement ou d'évolution représentables par deux polynômes en t à chacun desquels se superpose une composante aléatoire importante. Un nombre suffisant de différentiations fait disparaître le polynôme. Dans l'hypothèse d'indépendance pour les valeurs primitives des variables aléatoires, on démontre que le coefficient de corrélation entre deux différences d'ordre quelconque est le même que celui qu'on cherche entre les variables elles-mêmes.

(1) *Relations entre les indices économiques et le mouvement des mariages*, J. S. S. Paris, Mars 1911, pag. 80.

(2) *Etudes sur la formations et le mouvement des prix*, Paris, 1913.

Dans le même volume de « Biometrika », p. 269, on trouve un mémoire d'ANDERSON qui reprend et développe d'une façon approfondie l'idée émise par « Student ».

BEATRICE M. CAVE et K. PEARSON – « Biometrika », X, pp. 340-355.

ETHEL M. ELDERTON et K. PEARSON – « Biometrika » X, pp. 488-506 donnent des applications de ces méthodes nouvelles.

On trouve dans un travail de WARREN M. PERSONS – « Public. of the Am. Stat. Assoc. » Vol. XVI, 1917, p. 602, comme critique et idée nouvelle la remarque que les séries de variables aléatoires peuvent être à liaison interne.

En 1921, un important travail de G. UDNY YULE est exposé à la Royal Statistical Society, et suivi d'une intéressante discussion « J. R. S. S. » Vol LXXXIV, pp. 497-537.

On the time correlation problem, with special reference to the variate difference correlation method.

Ce travail approfondit et prolonge les idées préliminaires au travail de « Student », et signale un écueil possible de la méthode des différences, l'amplification démesurée de certaines oscillations de courtes périodes.

En réponse aux critiques de PERSONS et G. UDNY YULE, K. PEARSON et ETHEL M. ELDERTON reprennent la question d'une manière extrêmement large et suggestive. « Biometrika », XIV, pp. 281-310.

Utilisant un travail dû à EGON S. PEARSON « Biometrika », XIV, pp. 23-102 (spécialement pp. 37 et suivantes) ce mémoire fait progresser de façon substantielle l'étude des liaisons internes d'une série aléatoire.

Ces questions sont reprises par ANDERSON en 1923.

« Biometrika », XV, pp. 134-149.

En 1926 G. UDNY YULE consacre la « Presidential Address » de la Royal Statistical Society à une étude nouvelle du sujet:

Why do we sometimes get non sense correlations between time series. A study in sampling and the nature of time series.

« J. R. S. S. » Vol. LXXXIX, pp. 1-64.

Le début de ce mémoire est plutôt consacré dans notre terminologie à l'indice de linéarité, dans la section III, YULE propose une classification des séries aléatoires, par l'introduction des liaisons internes.

Enfin, ANDERSON a publié dans « Biometrika » en 1926 et 1927 un mémoire remarquable, riche d'idées, de résultats théoriques et d'applications.

Ueber die Anwendung der Differenzenmethode (*Variate difference method*) *bei Reihenausgleichungen, Stabilitätsuntersuchungen ud Korrelationsmessungen.*

« Biometrika », XVIII, pp. 293-320; XIX, pp. 53-86.

Ce travail traite comme on le voit, de la méthode des différences considérée comme un moyen d'ajuster des lois aux observations, d'étudier les liaisons internes des séries statistiques et les liaisons de ces séries entre elles.

17. *But du présent travail.* — Notre but n'est pas tant d'apporter des résultats pratiques nouveaux, que de mettre au point aussi complètement que possible, les idées théoriques et les résultats mathématiques qui nous semblent fondamentaux en cette matière. Dans la première partie, nous avons voulu dissiper la confusion que fait naître l'emploi du terme: coefficient de corrélation au lieu d'indice de linéarité.

Il reste maintenant diverses questions:

1) quels sont les moyens mathématiques les plus propres à la séparation des composantes. Nous n'avons rien à ajouter là-dessus au travail d'ANDERSON (1926-1927) dont nous parlons plus haut, qui montre en particulier que la méthode des différences conduit tout naturellement à une méthode régulière d'ajustement (pp. 300-301);

2) à la première question, assez formelle, on peut aiouter la suivante:

a-t-on des raisons de croire que la courbe de tendance comprenne en plus d'une courbe de croissance régulière, une courbe de fluctuations régulières, par exemple une fonction périodique superposée à un polynôme. Il est difficile de répondre si peu que ce soit, à une telle question. Si les idées que l'on possède sur le problème permettent une assimilation, même lointaine, avec un problème de dynamique, on peut penser que le système étudié possède une sorte de période propre de vibration autour d'une position d'équilibre dynamique lentement variable (1).

3) *Nous supposons qu'on emploie la méthode des différences.*

Il faut alors étudier l'effet de la différentiation sur une série de valeurs indépendantes d'une variable aléatoire. Quelles sont les propriétés de la série des différences. Quelle loi de probabilité suit

(1) Pour l'introduction de périodes approchés par une sommation de variables aléatoires, ce qui est un tout autre problème, voir: E. SLUTSKY («CR.» T. 185, 1927, p. 169).

un terme? Quel lien existe entre deux termes quelconques? Peut on retrouver complètement la loi de probabilité primitive connaissant la loi de la série des différences.

Pour deux séries, peut on de la loi de probabilité qui unit les différences, remonter entièrement à la loi de probabilité qui unit les variables?

Ces questions ont été étudiées jusqu'ici en se servant du seul coefficient de corrélation. Nous montrerons qu'elles sont susceptibles d'une solution complète, très simple.

4) On suppose la série des x pourvue de liaisons internes.

Quels sont les types les plus simples et les plus naturels de telles liaisons? Nous serons amenés d'abord à une étude dont l'intérêt propre est considérable, et nous retrouverons d'importants résultats sur l'affaiblissement des liens de probabilité obtenus par J. HADAMARD dans un ordre d'idées qui paraissait différent.

Nous verrons alors très clairement quels sont les effets de la différentiation sur ces séries enchaînées spéciales.

Ces résultats se généralisent immédiatement à deux séries ou à un nombre quelconque.

Les effets de la différentiation apparaissent de la façon la plus nette par l'emploi des fonctions caractéristiques.

Enfin, nous bornant au cas particulier des régressions linéaires nous retomberons sur les formules classiques de la théorie qui permettent de remonter des coefficients de corrélation des séries de différences à ceux des séries primitives.

18. *Les séries de variables aléatoires et les effets de la différentiation.* — Nous rappellerons rapidement la définition et les propriétés de la fonction caractéristique d'une loi de probabilité.

Nous supposerons qu'il s'agisse de variables continues, variant de a à b.

La fonction caractéristique est la valeur probable de l'exponentielle e^{itx}

$$\varphi(t) = E(e^{itx})$$

On sait que la loi de probabilité est déterminée par la connaissance de $\varphi(t)$; $f(x)$ étant la fonction des probabilités élémentaires:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t)\, e^{-itx}\, dt$$

La propriété importante de $\varphi(t)$ est la suivante:

La fonction caractéristique d'une somme de variables indépendantes est le produit des fonctions caractéristiques.

Il en résulte que, dans ce cas, les logarithmes des fonctions caractéristiques s'ajoutent pour donner le logarithme de la fonction résultante. Ce logarithme, que nous appellerons $\psi(t)$ est la fonction considérée par THIELE et qui fournit les semi-invariants, par son développement suivant les puissances de t.

Nous allons voir que la fonction ψ relative à la série des différences est reliée très simplement à la fonction ψ de la série primitive. Considérons d'abord la différence première.

Il est clair que sa fonction caractéristique est $\varphi(t)\ \ \varphi(-t)$, donc que la fonction ψ vaut:

$$\psi(t) + \psi(-t)$$

En général, si l'on écrit la différence d'ordre K sous la forme:

$$\alpha_0 \, x_{i+k} + \alpha_1 \, x_{i+k-1} + \ldots\ldots + \alpha_k \, x_i$$

les α sont les coefficients du developpement de $(1 - t)^k$. Il est clair alors que la fonction ψ relative à cette variable aléatoire est:

$$\psi(\alpha_0 t) + \psi(\alpha_1 t) \ldots + \psi(\alpha_k t)$$

Si ψ est développable en série entière, ou possède seulement un développement limité, les coefficients des puissances successives de t dans la fonction ψ résultante seront multipliés par

$$S_1^k \quad S_2^k \ldots S_m^k \ldots$$

S_m^k représentant la somme des puissances m des coefficients du binôme alterné.

Il est clair que si k est impair, les S_m^k seront nulles pour m impair. La fonction ψ résultante ne contient que des puissances paires de t, elle ne donnerait donc que les semi invariants d'ordre pair. Mais si k est pair, on obtiendra tous les semi-invariants de la fonction ψ cherchée en divisant par les nombres $S_2^k\ S_3^k \ldots$ les semi-invariants supposés connus de la fonction ψ relative à la différence d'ordre k.

19. — La notion de fonction caractéristique s'étend à un nombre quelconque de variables. On posera:

$\varphi(u_1\, u_2 \ldots u_n) = E\left[e^{i\,(u_1 x_1 + u_2 x_2 \ldots + u_n x_n)}\right]$ (i simbole de l'imaginaire).

La formule de FOURIER pour n variables, donne encore la fonction des probabilités élémentaires:

$$f(x_1\, x_2 \ldots x_n) = \frac{1}{(2\pi)^n} \underbrace{\int \ldots \int}_{n} \varphi(u_1\, u_2 \ldots u_n)\, e^{-i(u_1 x_1 \ldots + u_n x_n)}\, d_{u_1} d_{u_2} \ldots d_{u_n}$$

Nous appellerons encore ψ le logarithme de la fonction caractéristique.

Proposons nous maintenant de voir quelle est la structure interne de la série des différences. Les différentes valeurs sont liées par des lois de probabilité qu'il est facile de mettre en évidence. Considérons d'abord la série des différences premières.

Deux termes successifs sont liés. Deux termes qui ne se suivent pas sont indépendants. Pour obtenir la fonction caractéristique relative aux termes qui se suivent, formons la combinaison

$$\lambda = u(x_{i+i} - x_i) + v(x_{i+2} - x_{i+i})$$

L'espérance mathématique de $e^{i\lambda}$ est évidemment, $\varphi(t)$ étant la fonction caractéristique de la variable x:

$$\varphi(-u) \quad \varphi(u-v) \quad \varphi(v)$$

et la fonction ψ a la forme:

$$\psi(-u) + \psi(u-v) + \psi(v)$$

Si les différences sont d'ordre k, il est clair que deux termes séparés par de k–1 autres au plus sont liés. Au delà les termes sont indépendants.

Il est facile de trouver la liaison par la même méthode; soient les termes:

$$\alpha_0\, x_{i+k} + \ldots + \alpha_k\, x_i = m$$

$$\alpha_0\, x_{i+h+k} + \ldots + \alpha_k\, x_{i+h} = n$$

L'espérance mathématique de $e^{\,i(um+vn)}$ sera évidemment le produit d'un certain nombre de fonctions φ dont les premières ont des arguments de la forme $\alpha_l u$, les dernières des arguments $\alpha_l v$ celles du milieu étant des combinaisons linéaires de u, v, soit:

$$\varphi(\alpha_k\, u) \ldots \varphi(\alpha_{k-h-1}\, u)\, \varphi[\alpha_{k-h}\, u + \alpha_k\, v] \ldots \varphi[\alpha_0\, u + \alpha_h\, v] \ldots \varphi(\alpha_0\, v)$$

Si par exemple, on cherche les termes du second degré de la fonction ψ, on trouvera (en supposant que l'écart type de la variable x est pris pour unité):

$$-\frac{1}{2}\,[\,u^2\, S_2^k + v^2\, S_2^k + 2uv\,\Sigma\,]$$

Le coefficient de correlation ρ_k^h se calcule aisément.

$$\Sigma = \alpha_0 \alpha_h + \ldots + \alpha_{k-h} \alpha_k = (-1)^h C^{k-h}_{2k}$$

$$S_2^k = C_{2k}^k$$

$$\rho_h^k = \frac{(-1)^h C^{k-h}_{2k}}{C_{2k}^k} = \frac{(-1)^h (k!)^2}{(k-h)!\ (k+h)!}$$

Ce sont les résultats obtenus par YULE et ANDERSON.

20. *L'effet des différences sur une relation de probabilité entre séries.* — Considérons maintenant deux séries de valeurs, l'une relative à une variable aléatoire x, l'autre étant y:

$$x_1\ x_2 \ldots x_n$$
$$y_1\ y_2 \ldots y_n$$

Entre $x_i\ y_i$, existe un lien de probabilité, indépendant de i est que nous pouvons nous donner par sa fonction caractéristique à deux variables $\varphi(u\ v)$. Entre les différences premières $x_{i+1} - x_i$ $y_{i+1} - y_i$ existe un lien de probabilité dont la loi apparaît immédiatement. La fonction caractéristique, valeur probable de $e^{i\,[u\,(x_{i+1} - x_i) + v\,(y_{i+1} - y_i)]}$ est le produit de deux fonctions caractéristiques: $\varphi(uv)\ \varphi[-u\ -v]$.

Il est clair que les différences d'ordre K aient pour fonction caractéristique $\varphi(\alpha_0\, u, \alpha_0\, v) \ldots \varphi(\alpha_k\, u, \alpha_k\, v)$
et par conséquent la fonction ψ a pour valeur

$$\sum_{i=0}^{k} \psi\,[\alpha_i\, u, \alpha_i\, v]$$

Ce résultat extrêmement simple permet de passer de la loi primitive à la loi des différences, et inversement. Les groupes homogènes en $u\ v$ des degrés $2, 3 \ldots m, \ldots$ sont en effet dans le rapport

$$S_2^k\ S_3^k \ldots S_m^k$$

Par conséquent une différence d'ordre impair détermine tous les groupes d'ordre pair, une différence d'ordre pair détermine tous les groupes sans exception.

Si l'on se borne au groupe de second ordre, dont les coefficients sont les trois moments du second ordre, on retrouve le résultat de « Student ».

Le coefficient de corrélation conserve sa valeur dans les différences successives. On voit qu'on peut déduire dans le cas général, de façon complète et simple, la loi de probabilité primitive de la loi de probabilité des différences.

Cas d'un nombre quelconque de séries. — On peut avoir à comparer non pas deux séries de variables aléatoires, mais trois ou plus. Si l'on conserve l'hypothèse de l'indépendance à l'intérieur de chaque série, et d'une liaison de probabilité à n variables unissant seulement les groupes des termes de même indice, l'inconnue du problème est une loi de probabilité à n variables, au lieu d'être à deux variables.

Il n'y a rien à modifier aux résultats précédents. On peut, par les mêmes moyens, déduire tous les groupes homogènes de l'une des fonctions ψ des groupes homogènes de l'autre, en multipliant ou divisant par les facteurs S_m^k introduits auparavant.

21. *Les séries à liaison interne – Etude préliminaire des lois de probabilité liées.* — Nous avons besoin d'approfondir un peu le concept de loi de probabilité à plusieurs variables. Supposons d'abord qu'il y en ait deux. On peut se donner la loi de probabilité élémentaire, soit: $F\ (x\ y)\ dx\ dy$.

Nous savons qu'il lui correspond une fonction caractéristique $\Phi\ (u\ v)$ qui la détermine entièrement.

Mais on peut opérer autrement. La loi de probabilité de x seul ayant pour loi élémentaire $A\ (x)\ dx$, donnons nous la loi élémentaire de y, x étant fixé. Soit: $f_x\ (y)\ dy$, on aura:

$$F\,(xy) = A\,(x)\,f_x\,(y)$$

$f_x\,(y)\,dy$ donnera ce que nous appellerons la loi liée (1) pour y.

On peut évidemment considérer une loi liée pour x et l'on aura

$$F\,(xy) = B\,(y)\;g_y\;(x)\,.$$

Ces lois liées sont souvent plus commodes à considérer que la loi à 2 variables. Par exemple, le rapport de corrélation de PEARSON fait intervenir la loi liée.

Il existe évidemment une fonction caractéristique liée. Soit $\varphi_x\,(v)$

Puisque l'on a:

$$\Phi\,(uv) = \int\!\!\int e^{i(ux+vy)}\,A\,(x)\,f_x\,(y)\;dx\,dy = \int e^{iux}\,A\,(x)\;\varphi_x\,(v)\;dx$$

on aura, par l'inversion:

$$A(x)\;\;\varphi_x\,(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi\,(uv)\;e^{-iux}\;du$$

(1) TCHOUPROW. *Grundbegriffe und Grundprobleme der Korrelationstheorie.*

La première de ces équations permet le calcul de $\Phi(uv)$ connaissant la loi de x et la fonction caractéristique liée. La deuxième fournit la loi liée connaissant la fonction caractéristique générale. Il suffit au second membre, de développer $\Phi(uv)$ suivant les puissances de v pour obtenir tous les moments de la loi liée en fonction de séries de moments de la loi générale.

$$A(x) = \frac{1}{2\pi}\int \alpha_o(u)\, e^{-iux} du \qquad A(x)\, M_{x/p} = \frac{1}{2\pi}\int \alpha_p(u)\, e^{-iux} du$$

Chaque moment lié dépend donc d'une infinité de moments de la loi générale, $\alpha_p(u)$ a en effet pour coefficients tous les moments

$$E\,(x^\lambda y^p)$$

où p est fixe, λ allant de o à $+\infty$ par valeurs entières.

Si par exemple, on veut l'équation de la ligne de régression de y en x, on aura:

$$M_{x/1} = \frac{\int \alpha_1(u)\, e^{-iux}\, du}{\int \alpha_0(u)\, e^{-iux}\, du}$$

22. *Lès lois à* n *variables.* — Considérons trois variables aléatoires, liées par une loi de probabilité. On peut évidemment soit se donner la probabilité élémentaire:

$$F(x_1\, x_2\, x_3)\; dx_1\, dx_2\, dx_3$$

soit se donner la loi pour x_1, $A(x_1)\, dx_1$, la loi pour x_2 quand x_1 est fixé, $f_{x_1}(x_2)\, dx_2$, la loi de probabilité pour x_3 quand x_1 et x_2 sont fixés, $g_{x_1 x_2}(x_3)\; dx_3$. On aurait:

$$F(x_1\, x_2\, x_3) = A(x_1)\, f_{x_1}(x_2)\, g_{x_1\, x_2}(x_3)$$

On peut aussi se donner $A(x_1)\; dx_1$ et la loi de probabilité liée en x_1 des deux variables $x_2\, x_3$ soit $h_{x_1}(x_2\, x_3)\, dx_2\, dx_3$

Il est clair que, suivant le problème à étudier, l'une ou l'autre de ces formes sera plus indiquée.

Si les variables sont indépendantes, les lois liées ne dépendent pas des variables de liaison, et la fonction caractéristique de la loi de probabilité à trois variables est le produit de trois fonctions.

D'autres cas importants et plus généraux peuvent se présenter. $g_{x_1\, x_2}(x_3)$ peut par exemple ne pas dépendre de la variable x_1, tout

en continuant à dépendre de la variable x_2. Connaître x_2 est une liaison pour x_3, mais connaître en plus x_1 n'y ajoute rien. Nous allons appliquer ces remarques, évidemment valables pour un nombre quelconque de variables.

23. *Les séries aléatoires régulièrement enchaînées.* — Etant donnée une série aléatoire à liaisons internes, c'est-à-dire telle qu'une des variables ne soit pas indépendante de toutes les autres, l'hypothèse la plus générale est que ces n variables sont liées par une loi de probabilité qui ne soit astreinte à aucune restriction. Dans les cas qui nous occupent, cette généralité serait excessive, à la fois peu propre à une explication, que nous cherchons, et à une vérification expérimentale.

Il nous semble que si, par exemple, les valeurs observées sont annuelles, et qu'elles ne soient pas indépendantes, l'hypothèse la plus simple à faire est que la loi de probabilité d'une variable est liée si l'on connaît les valeurs qui l'ont précédée, mais que cette liaison ne doit comporter qu'un petit nombre des valeurs précédentes.

La probabilité d'une valeur x_i peut dépendre de x_{i-1} seulement, ou de x_{i-1} et x_{i-2} seulement. D'autre part, ces liaisons par petits groupes doivent se répeter tout le long de la série.

Nous allons d'abord poursuivre les conséquences de cette idée très simple dans le cas où la loi de probabilité liée d'un terme ne dépend que du terme précédent. Nous pourrons dire, adoptant la terminologie de MARKOFF, que les valeurs considérées forment une chaîne, nous dirons que cette chaîne est régulière et procède par deux pour exprimer que la loi de probabilité liée pour x_{i+1} est de la forme:

$$f(x_i\,,\ x_{i+1})\ dx_{i+1}$$

la fonction f étant la même, quel que soit l'indice i.
soit dans ce cas $A(x_1)\,dx_1$ la loi de probabilité de x_1.
la loi complète pour $x_1\ x_2$ sera donc:

$$A(x_1)\ f(x_1\ x_2)\,dx_1\,dx_2$$

Nous devons faire l'hypothèse que la loi de probabilité d'une variable d'indice quelconque, considérée *a priori*, est la même que celle de x_1. Il en résulte immédiatement la condition:

$$\int_a^b A(x_1)\,f(x_1\ x_2)\,dx_1 = A(x_2)$$

a et b étant les extrémités de l'intervalle où tombe la variable aléatoire x. D'autre part, nous avons la condition évidente:

$$\int_a^b f(x_1 \, x_2) \, dx_2 = 1$$

indépendamment de x_1

On voit qu'on est amené de suite à considérer $f(x_1 \, x_2)$. comme le noyau d'une équation intégrale ou plutôt d'un couple d'équations intégrales associées, possédant une constante caractéristique égale à l'unité, les solutions correspondantes de l'équation homogène étant A (x) et l'unité (1).

Nous allons retrouver dans un instant cette équation intégrale.

24. *Les liaisons par intermédiaires.* — Considérons les deux variables aléatoires $x_1 \, x_3$. Nous avons vu que, par hypothèse, si x_2 est fixé elles sont indépendantes. Mais si x_2 n'est pas fixé, elles ont entre elles un lien de probabilité qu'il est facile de trouver. La loi de probabilité de x_3 liée par x_1 est évidemment:

$$f^{(2)}(x_1 \, x_3) \, dx_3 = dx_3 \int_a^b f(x_1 \, x_2) \, f(x_2 \, x_3) \, dx_2$$

La fonction $f^{(2)}$ résulte donc d'une itération de la fonction f.

Il est intuitif et de vérification immédiate que le noyau $f^{(2)}$ a la constante caractéristique 1, avec les mêmes solutions A (x) et l'unité.

Il est clair que la liaison de $x_1 \, x_{p+1}$ est représentée par la fonction $f^{(p)}(x_1 \, x_{p+1})$ qui résulte de $p-1$ itérations successives.

Il est d'autre part évident qu'une telle liaison peut être conçue comme résultant par exemple de la liaison entre x_1 et x_i, suivie de la liaison entre x_i, x_{p+1}, ce qui donne:

$$f^{(p)}(xy) = \int_a^b f^q(xs) \, f^r(s\,,y) \, ds \qquad q + r = p$$

en particulier:

$$f^{(p)}(xy) = \int_a^b f^{(p-1)}(xs) \, f(sy) \, ds = \int_a^b f(xs) \, f^{(p-1)}(sy) \, ds$$

Il paraît naturel que ces liaisons s'affaiblissent avec la distance des termes; autrement dit que la connaissance de x ait peu d'influence sur la loi de probabilité d'un terme éloigné y.

(1) Voir par ex. GOURSAT. *Traité d'analyse*. Tome III.

D'autre part, la loi liée agissant de moins en moins, nous devons compter, si p augmente indéfiniment, que cette loi liée doit avoir pour limite la loi suivie par la variable quand x est quelconque, c'est–à–dire la loi A (y). Ainsi, nous sommes amenés à penser que $f^{(p)}(xy)$ quand p augmente indéfiniment à une limite indépendante de x, et que cette limite est A (y). Ces résultats, auxquels nous sommes amenés de la manière la plus naturelle, ont été établis récemment par J. HADAMARD et HOSTINSKI (1), à propos de questions voisines, mais dans un ordre d'idées différent, inauguré par HENRI POINCARÉ.

Nous renverrons pour la démonstration aux travaux de ces, auteurs en particulier à la note de J. HADAMARD (CR. T. 186, 1928 p. 189).

Le principe en est très simple. Les fonctions $f^{(p)}(xy)$ apparaissent comme des moyennes (2), leurs limites extrêmes vont donc en se resserrant à mesure que p augmente. D'autre part, dès l'instant où l'on a démontré l'existence de cette limite, elle doit, d'après l'équation

$$f^{(p)}(xy) = \int f^{(p-1)}(xs)\, f(sy)\, ds$$

être une solution de l'équation intégrale:

$$\Phi(xy) = \int_a^b \Phi(xs)\, f(sy)\, ds$$

On démontre aisément que la constante caractéristique 1 est simple, et comme nous savons que cette équation possède la solution A (y), la limite coïncide nécessairement avec cette fonction.

25. *Cas particulier des régressions linéaire.* — Ce théorème général prend une forme très simple dans la cas ou la régression de x_{i+1} en x_i est linéaire, c'est–à–dire:

$$\int_a^b f(x_1\, x_2)\, x_2\, dx_2 = r\, x_1$$

(1) J. HADAMARD « CR. Ac. Sces » T. 185–1927, p. 5, T. 186, 1928 p. 189 – Congrès de Bologne, 1928. — B. HOSTINSKY. « CR. » T. 186, 1928, p. 59, 186, 1928, p. 487 – Congrès de Bologne, 1928.

(2) F. M. URBAN, *Grundlagen der Wahrscheinlichkeitsrechnung-:.* . Berlin, 1923, pp. 166-171.

On sait que, dans ce cas, le coefficient r coïncide avec le rapport de corrélation de y en x et donne une mesure de la liaison entre x et y. On a alors évidemment

$$\int_a^b f^{(2)}(x_1\, x_3)\, x_3\, dx_3 = \int\int f(x_1\, x_2)\, f(x_2\, x_3)\, x_3\, dx_2\, dx_3 = \int_a^b f(x_1\, x_2)\, r\, x_2\, dx_2$$

Cette valeur est donc $r^2 x_1$. D'une façon générale:

$$\int_a^b f^{(p)}(xy)\, y\, dy = r^p x$$

Le coefficient de corrélation s'affaiblit donc en progression géométrique.

26. *La liaison par trois.* — La chaîne la plus simple qu'on puisse considérer ensuite est celle où la loi de probabilité liée d'un couple x_{i+1}, x_{i+2} ne dépend, parmi toutes les valeurs précédentes, que de x_i. Nous noterons cette loi liée

$$F\ (x_i\, x_{i+1}\ x_{i+2})\ dx_{i+1}\ dx_{i+2}$$

Il est alors facile de trouver, par exemple, la loi liée du couple $x_4\, x_5$ en fonction de x_1. Ce sera:
$f^{(2)}\,[\,x_1\, x_4\, x_5\,]\ \ dx_4\ dx_5$ en posant:

$$f^{(2)}[x_1\, x_4\, x_5] = \int\int f(x_1\, x_2\, x_3)\, f(x_3\, x_4\, x_5)\, dx_2\, dx_3$$

On voit qu'on est amené à une itération d'une nature particulière.

Cette liaison par trois donne lieu à des développements analogues à ceux que nous avons obtenus pour la liaison par deux. Sans insister sur ce point, nous allons seulement traiter le cas où les moyennes x_{i+1}, x_{i+2} sont linéaires en x_i. On aura donc:

$$\int\int f(x_1\, x_2\, x_3)\, x_2\, dx_2\, dx_3 = r\, x_1$$
$$\int\int f\, (x_1\, x_2\, x_3)\, x_3\, dx_1\, dx_3 = r' x_1$$

Les liaisons plus lointaines obéissent alors à une loi simple.

La moyenne de x_4, par exemple, est proportionnelle à x, avec le coefficient rr', la moyenne de x_5 avec le coefficient r'^2.

D'une façon générale, la suite des coefficients de corrélation est la suivante:

$$1,\ r,\ r',\ rr',\ r'^2\ rr'^2\ r'^3 \ldots$$

L'affaiblissement des corrélations est encore ici très simple.

Il est d'ailleurs évident que ces résultats s'étendent à la liaison par quatre ... etc. ... On aurait par la suite des coefficients:

$$1 \quad r \quad r' \quad r'' \quad rr'' \quad r'r'' \quad r''^2 \quad rr''^2 \quad r'r''^2 \quad r''^3 \ldots$$

Nous désignerons d'une façon générale cette suite par

$$r_0 \; r_1 \; r_3 \; r_2 \ldots r_k \ldots$$

Sa structure dépendra d'après ce qui précède du mode de liaison de la chaîne.

27. *Les liaisons de la série des différences dans le cas d'une série quelconque.* — Considérons une série de n valeurs d'une même variable aléatoire, ces n valeurs possédant une liaison de probabilité quelconque. Cette liaison sera donnée par une fonction des probabilités élémentaires à n variables ou, ce qui revient au même, par sa fonction caractéristique.

Il est bien clair que, de la connaissance de cette loi générale, on peut déduire la loi de probabilité commune à toutes les variables aléatoires, la loi de probabilité d'un nombre quelconque de ces variables, en particulier, de deux variables qui se suivent, de trois variables, d'un nombre quelconque de variables qui se suivent.

Pour étudier les effets d'une seule différentiation, on devra chercher pour cette série des différences, la loi de probabilité d'un terme quelconque et les lois de probabilité qui lient ces termes entre eux. On peut trouver individuellement ces diverses lois.

Par exemple, quelle loi suit la différence $x_{i+1} - x_i$. Il suffit de se reporter à la fonction caractéristique de la loi du couple, doit $\Phi(\alpha\beta)$. La fonction caractéristique cherchée sera évidemment $\Phi(-t, t)$. Elle dépendrait généralement de x_i, x_{i+1}, mais dans le cas que nous étudions, elle devra en être indépendante.

La différence seconde $x_{i+2} - 2x_{i+1} + x_i$ suit une loi qu'on déduit immédiatement de la loi des trois termes successifs. Si $\Phi(\alpha\beta\gamma)$ est la fonction caractéristique, la fonction cherchée est:

$$\Phi(t, -2t, t)$$

De même la loi qui lie deux différences premières fait intervenir trois termes consécutifs. Elle se déduit donc de $\Phi(\alpha\beta\gamma)$. On trouve immédiatement:

$$\Phi(-u, u-v, v)$$

Nous allons appliquer ces remarques générales au cas spécial des chaînes à regréssion linéaire, où il est aisé de remonter des coefficients de corrélation pour les différences aux coefficients primitifs.

28. *Chaîne à liaisons par deux, trois...* — Il est facile d'obtenir par plusieurs procédés qui doivent concorder, le coefficient de corrélation de la liaison par deux, trois . . . , en utilisant la suite des différences.

Dans tout ce qui va suivre, en effet, nous n'aurons affaire qu'au groupe de termes du second degré de la fonction caractéristique.

Imaginons que nous puissions opérer sur les différences premières.

Nous pouvons étudier la loi de probabilité d'un terme quelconque et les liaisons des termes entre eux.

Le résultat s'obtient d'un coup en passant de la fonction caractéristique de la série des x_i à la fonction caractéristique de la série des différences. Si $\Phi\,(u_1\, u_2\, u_2 \ldots u_n)$ est la première, la deuxième est évidemment:

$$\Phi\,[-v_1\,,\; v_1 - v_2\;\; v_2 - v_3 \ldots v_{n-2} - v_{n-1}\,,\; v_{n-1}]$$

Cette méthode s'applique immédiatement à la série des différences d'ordre quelconque k. On déduit de la fonction Φ à n variables la fonction à n–k variables des différences. Avec les notations du § 18, on écrira:.

$$\Delta^k x_i = \alpha_0\, x_{i+k} + \ldots + \alpha_k\, x_i$$

Par conséquent, on voit de suite que la fonction inconnue est:

$$\Phi\,[\alpha_k\, v_1\;\; \alpha_{k-1}\, v_1 + \alpha_k\, v_2\;\; \alpha_{k-2} v_1 + \alpha_{k-1}\, v_2 + \alpha_k\, v_3 \ldots \alpha_0\, v_{n-k}] \,..$$

Le terme du second degré de la fonction Φ est proportionnel à:

$$r_0\,[u_1^2 + u_2^2 + u_n^2] + 2r_1\,[u_1\, u_2 \ldots + u_{n-1}\, u_n] + \ldots + 2r_{n-1}\, u_1\, u_n$$

Il suffit de substituer les expressions écrites plus haut pour obtenir d'un coup les fluctuations des différences et leurs liaisons entre elles à une distance quelconque. En utilisant les valeurs de Σ, S_2^k du § 19, on trouve alors les valeurs générales rencontrées d'abord par EGON S. PEARSON (Thèse, *Biometrika*, XIV, pp. 37 et suivantes) puis par ANDERSON (*Biometrika*, XV, p. 136) qui leur a donné une forme élégante et brève.

Si l'on conçoit la suite:

$$r_0\; r_1\; r_2 \ldots r_k \ldots$$

prolongée à partir de r_0 par la suite symétrique, et qu'on désigne par $\Delta^k r_i$ la combinaison linéaire des r avec les coefficients $\alpha_0\, \alpha_1 \ldots$ on trouvera, pour la fluctuation de la différence k^{me} des x:

$$\sigma^2 = \Delta^{2k} r_k$$

Par exemple, pour les différences secondes:

$$\sigma^2 = 6r_0 - 8r_1 + 2r_2$$

On trouvera de même pour le coefficient de $v_1\, v_{1+h}$ dans la fonction caractéristique l'expression

$$\Delta^{2k} r_{k+h}$$

Par exemple, pour les différences secondes, on trouvera pour $h = 1$:

$$-4r_0 + 7r_1 - 4r_2 + r_3$$

Dans le cas de la liaison par deux, il n'y a que deux inconnues la fluctuation de x et le coefficient de corrélation r.

Il n'y a aucune difficulté à en déterminer diverses valeurs, qui devront concorder. On pourra utiliser les indications données par K. Pearson et Ethel M. Elderton (*Biometrika*, 14, pp. 282-310) et la table dressée par J. Anderson (p. 310).

La liaison par trois introduirait 3 inconnues $\sigma_x\, r\, r'$.

Si l'on imagine que le mouvement séculaire soit convenablement éliminé par les différences secondes, on connaîtra d'abord la valeur de

$$\sigma_x^2\,(6 - 8r + 2r')$$
$$\sigma_x^2\,(-4 + 7r - 4r' + rr')$$
$$\sigma_x^2\,(1 - 4r + 6r' - 4rr' + r'^2)$$

Les autres équations pour des valeurs plus grandes de h devront être compatibles avec celles là.

Il est bien évident que si aucune hypothèse particulière n'est introduite pour les coefficients r_k, leur détermination ne pourra résulter de la seule considération de la suite des différences. Mais il ne saurait guère arriver qu'on soit si complètement ignorant du problème que l'on étudie. Sinon, c'est que ce problème ne serait pas mûr pour un essai de théorie.

29. *La relation entre deux séries.* — Plaçons nous maintenant en présence de deux séries, l'une de valeurs d'une variable aléatoire x, l'autre série correspondant à la variable y. Deux termes entre les-

quels nous aurons établi une correspondance en leur donnant le même indice seront liés par une loi de probabilité que nous voulons rechercher, nous supposons les deux séries à liaisons internes. De plus, nous supposons que la série double des couples x_i y_i est elle aussi à liaison internes. La loi de probabilité du couple x_i y_i variera si l'on connaît par exemple le couple précédent. Si elle ne varie qu'avec le couple précédent; ou d'une façon plus précise, si la loi de probabilité du couple x_i y_i , liée par la connaissance de tous les couples qui le précédent ne dépend que du couple immédiatement antérieur, nous dirons que nous avons une double chaîne à liaison par deux, elle sera régulière si la loi de probabilité liée est partout la même. C'est la généralisation naturelle des notions précédentes. Considérons le couple x_1 y_1. La loi de probabilité élémentaire sera $A\ (x_1\ \ y_1)\ dx_1\ \ dy_1$

Si $F\ (x_1\, y_1\ x_2\, y_2)\ dx_2\, dy_2$ est la loi liée, nous avons pour la loi à 4 variables.

$$A\ (x_1\, y_1)\ F\ (x_1\, y_1\ ,\ x_2\, y_2)\ dx_1\, dx_1\, dx_2\ \ dy_2$$

Nous imposerons la condition que la loi de x_2 y_2 quand x_1 y^1 sont quelconques soit la même que celle de x_1 y_1.

Nous avons la condition:

$$A\ (x_2\, y_2) = \iint F\ (x_1\, y_1\, x_2\, y_2)\ A\ (x_1\, y_1)\ dx_1\, dy_1$$

à laquelle on peut ajouter

$$\iint F\ (x_1\, y_1\, x_2\, y_2)\ dx_2\, dy_2 = 1$$

quels que soient x_1 y_1.

Ces conditions sont tout à fait analogues à celles du § 23. La fonction F est le noyau de deux équations intégrales associées ayant une constante caractéristique x égale à l'unité, avec les solutions $A\ (x_1\, y)$ et l'unité pour les équations homogènes.

30. *Les liaisons par intermédiaires.* — Il est évident que la liaison de $x_3\, y_3$ et de $x_1\, y_1$ s'obtiendra par une itération du noyau F. On pourra écrire:

$$F^{(2)}\ (x_1\, y_1\, x_3\, y_3) = \iint F\ (x_1\, y_1\, x_2\, y_2)\ F\ (x_2\, y_2\, x_3\, y_3)\ dx_2\, dy_2$$

Toute la théorie du § 14 s'applique. Les liaisons s'affaiblissent avec la distance. Quand le nombre des termes intermédiaires augmente indéfiniment, la loi liée tend vers la loi $A\ (xy)$.

La démonstration rigoureuse de ces résultats assez intuitifs suit exactement la même marche.

Bien entendu, ces résultats s'étendent à une liaison par trois etc.

31. *Les liaisons de la double chaîne des différences.* — L'étude théorique des liaisons de la double chaîne des différences se fait aisément suivant la méthode du § 28. Il suffit de considérer les $2\,n$ variables aléatoires $x_i\ y_i$ et leur fonction caractéristique, fonction de $2\,n$ variables $u_i\,v_i$:

$$\Phi\,(u_1\ u_2\ .\ .\ .\ u_n\ \ v_1\ v_2\ .\ .\ .\ v_n)$$

Cette fonction Φ nous donne toutes les liaisons que nous pouvons désirer connaître. Si tous les v_i sont nuls, on trouve la fonction caractéristique de la série des x.

On en tirerait aisément, par les méthodes du § 21 toutes les lois liées.

Considérons maintenant la double chaîne des différences $\Delta^k\,x_i$ $\Delta^k\,y_i$ qui comporte n–k couples:

$$X_1\,X_2\,.\,.\,.\,X_{n-k}$$
$$Y_1\,Y_2\,.\,.\,.\,Y_{n-k}$$

On peut obtenir immédiatement la fonction caractéristique relative à ces $2\,(n–k)$ variables aléatoires. Il suffit en effet de former la combinaison linéaire :

$$U_1\,X_1 + \ldots + U_{n-k}\,X_{n-k} + V_1\,Y_1\,\ldots\,V_{n-k}\,Y_{n-k}$$

C'est une forme linéaire des $x_i\ y_i$, ses coefficients ont déjà été rencontrés au § 28.

Leur forme générale est d'ailleurs très simple:

$$\alpha_0\,U_i + \alpha_1\,U_{i+1}\,.\,.\,.\,+ \alpha_k\,U_{i+k} = (-1)^k\,\Delta^k\,U_i$$

Ainsi le terme général, *quand il comprend tous les termes est* $(-1)^k\,\Delta^k\,U_i$, les termes des extrémités rentrent dans cette expression générale en convenant de prolonger par des zéros aux extrémités la série des U_i . Cette remarque générale étant faite, il suffit de remplacer, dans la fonction Φ , u_i par $(-1)^k\,\Delta^k\,U_i$, v_i par $(-1)^k\,\Delta^k\,V_i$ pour obtenir la nouvelle fonction caractéristique.

Le problème posé par la méthode des différences est donc de remonter de cette dernière fonction à la fonction primitive.

On voit qu'il est plus difficile que celui du § 20 où il s'agissait par la connaissance d'une fonction de deux variables, de remonter à la connaissance d'une autre fonction de deux variables.

Dans le cas présent, nous aurons moins d'équations que d'inconnues, si aucune restriction n'est apportée à la loi de probabilité générale.

Mais, comme nous l'avons déjà dit, une hypothèse aussi générale signifierait une bien complète ignorance du problème que nous désirons traiter.

Les liaisons par petits groupes, dont nous avons parlé, semblent être dans bien des cas l'hypothèse naturelle.

D'autres idées peuvent être suggérées par chaque question particulière, elles apporteront des restrictions à la généralité de la fonction caractéristique et peuvent permettre de la déterminer complètement par la fonction analogue relative aux différences.

32. — Pour obtenir des conclusions simples, nous supposerons encore les régressions linéaires, de sorte que la détermination des coefficients de corrélation suffise à donner une idée assez exacte des lois de probabilité.

Prenons d'abord le cas où la double chaîne présente la liaison par deux, la loi liée de $x_{i+1}\ y_{i+1}$ ne dépendant que du couple antérieur. Cette hypothèse suffit à assurer la liaison par deux des chaînes des x et des y. Soient r et s les coefficients de corrélation. On sait qu'alors la corrélation des termes éloignés a des coefficients décroissant en progression géométrique.

Considérons alors les moyennes liées de $x_{i+1}\ y_{i+1}$. Supposons que ce soient des fonctions linéaires de $x_i\ y_i$

$$ax_i + by_i$$
$$cx_i + dy_i$$

Il est clair que le couple $x_{i+2}\ y_{i+2}$ considéré comme lié à $x_i\ y_i$ aura ses moyennes linéaires, données par l'itération de la transformation linéaire écrite plus haut, soit:

$$(a^2 + bc)\, x_i + (ab + bd)\, y_i$$
$$(ae + cd)\, x_i + (d^2 + bc) y_i$$

Appellons maintenant ρ_0 le coefficient de corrélation de $x_i\, y_i$, ρ_1 celui de $y_i\, x_{i+1}$, ρ_{-1} celui de $x_i\, y_{i+1}$, on aura:

$$\begin{cases} r = a + b\rho_0 \\ \rho_1 = a\rho_0 + b \\ \rho_{-1} = c + d\rho_0 \\ S = c\rho_0 + d \end{cases}$$

En opérant de même avec la substitution itérée, on trouve:

$$\begin{cases} r_2 = ar + b\rho_{-1} \\ \rho_2 = a\rho_1 + bs \\ \rho_{-2} = cr + d\rho_{-1} \\ s_2 = c\rho_1 + ds \end{cases}$$

Si l'on veut que $r_2\, s_2$ soient les carrés de r, s, on trouve les conditions:

$$(r\rho_0 - \rho_1)\ (r\rho_0 - \rho_{-1}) = 0$$
$$(s\rho_0 - \rho_1)\ (s\rho_0 - \rho_{-1}) = 0$$

Ce système se décompose en deux:

$$\begin{cases} \rho_1 = r\rho_0 \\ \rho_{-1} = s\rho_0 \end{cases} \qquad \begin{cases} \rho_1 = s\rho_0 \\ \rho_{-1} = r\rho_0 \end{cases}$$

Nous adopterons l'un ou l'autre de ces systèmes de conditions. On voit aisément qu'ils ont pour conséquences:

$$\begin{cases} \rho_p = \rho_0\, r^p \\ \rho_{-p} = \rho_0\, s^p \end{cases} \qquad \begin{cases} \rho_p = \rho_0\, s^p \\ \rho^{-p} = \rho_0\, r^p \end{cases}$$

Quelle est leur signification? (1)
Imaginons les quatre valeurs:

$$\begin{matrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{matrix}$$

Les coefficients r s correspondent aux lignes, le coefficient ρ_0 aux deux colonnes, les coefficients $\rho_1\, \rho_{-1}$ aux diagonales.

(1) Cfr. Karl Pearson et Ethel M. Elderton. « Biometrika », 14, p. 297. Voir aussi les travaux de K. Pearson sur l'hérédité ancestrale. « Philos. Transac. », A. 1886. « Proceed. of the Royal Soc. », B. 81 – 1909.

Si l'on va de x_i à y_{i+1}, par l'intermédiaire de y_i, on est amené à considérer la loi de y_{i+1} liée par la connaissance de couple $x_i\, y$. Si cette loi liée est indépendante de xi, on aura:

$$\rho^{-1} = s\rho_0$$

Si au contraire, on passe par x_{i+1} et que la loi de y_{i+1}, liée par $x_i\, x_{i+1}$ ne dépende pas de x_i, on aura:

$$\rho_{-1} = r\rho_0$$

Telle est la signification générale non nécessaire, qu'on peut donner aux conditions obtenues.

On voit que, sous les hypothèses habituelles de régressions linéaires, l'un ou l'autre de ces systèmes des conditions est nécessaire.

33. — Si d'une façon générale, nous appelons seulement $r_k\, s_k\, \rho_p\, \rho_{-p}$ les différents coefficients de corrélation des éléments de la double chaîne, il est facile de trouver les coefficients de corrélation pour les différences. Dans ce cas, en effet, nous opérerons seulement sur la partie du second degré de la fonction caractéristique, à laquelle nous appliquons la proposition générale obtenue plus haut.

Dans ces conditions, on obtient très aisément les résultats de EGON S. PEARSON et ANDERSON; l'espérance mathématique du produit $\Delta^k\, x_i\, \Delta^k\, y_{i+h}$ a pour valeur $\Delta^{2k}\, \rho_{k+h}$ (Bien entendu, nous supposerons $x_i\, y_i$ mesurés en écarts types).

Ce résultat permet dans le cas des chaînes régulières, de remonter aux coefficients de corrélation primitifs.

Un exemple suffira. Supposons que la double chaîne soit à liaison par deux, et qu'il suffise de prendre les différences secondes pour éliminer la tendance. La liaison entre termes de même rang donnera dans la fonction caractéristique un groupe du second degré ayant la forme suivante:

$$\alpha^2\,[6r_0 - 4r_1 + 2r_2] + \beta^2\,[6s_0 - 4s_1 + 2s_2]$$
$$+ 2\alpha\beta\,[\rho_2 - 4\rho_1 + 6\rho_0 - 4\rho_{-1} + \rho_{-2}]$$

En introduisant les hypothèses faites:

$$r_1 = r \qquad r_2 = r^2 \qquad s_1 = s \qquad s_2 = s^2$$
$$\rho_1 = \rho_0\, s \qquad \rho_2 = \rho_0\, s^2 \qquad \rho_{-1} = \rho_0\, r \qquad \rho^{-2} = \rho_0\, r^2$$

On voit alors immédiatement que les valeurs de r, s étant déterminées par l'étude isolée che chacune des séries, la valeur de ρ_0 sera

connue par la corrélation des termes correspondants des séries de différences secondes. Les relations entre termes plus éloignés devront donner pour ρ_0 des valeurs concordantes.

Résumé et conclusions. — Nous avons considéré un ou plusieurs phénomènes se développant dans le temps, et variant avec lui. On dit parfois que le but poursuivi est de chercher la corrélation entre l'un des phénomènes et le temps, ou entre les deux phénomènes. L'emploi de ce mot nous paraît à éviter. Il peut être trop expressif et prête à confusion. Pour ne pas créer un mot nouveau, on pourrait peut-être restreindre à ce cas l'emploi du mot covariation, employé systématiquement par L. MARCH.

Dire que deux grandeurs sont en covariation est l'affirmation purement formelle que l'on les voit varier ensemble.

Ceci posé, on peut ou bien vouloir décrire une covariation pour des fins pratiques, ou bien avoir des raisons d'expliquer cette covariation par un lien véritable.

Ce n'est que dans le deuxième cas qu'on peut prétendre à obtenir une loi. On se trouve alors placé devant le problème de l'induction; il faut renforcer le plus possible, par l'observation et le raisonnement, les raisons *a priori* que l'on a de croire à une relation.

Si l'on est au contraire dans le premier cas, aucune description mathématique, si précise soit elle, ne possède le pouvoir de nous pousser vers une induction qui manque de bases raisonnables.

La qualité de notre description, son aptitude à remplacer l'ensemble des observations, peuvent être cotées à l'aide d'indices de linéarité dont le premier, le plus simple, est en somme un coefficient de covariation linéaire, mais non un coefficient de corrélation.

Dans une deuxième partie de ce travail, nous avons porté notre attention sur la composante aléatoire des phénomènes qui varient dans le temps, et complété les résultats obtenus dans cette voie par la méthode des différences.

Il s'agit alors de déterminer des lois de probabilité.

Nous avons montré que, dans le cas le plus général, si l'on considère un nombre quelconque de séries aléatoires, et qu'on agisse sur elles par la méthode des différences, les liaisons de probabilités des séries de différences se déduisent par une règle extrêmement simple des liaisons des séries primitives.

Dans la méthode des différences, il faut faire l'opération inverse; nous avons montré que si les variables sont indépendantes dans

chaque série, et qu'il n'existe de lien qu'entre des variables de même indice de toutes les séries, on peut obtenir complètement la loi de probabilité primitive connaissant la loi des différences. Ce résultat n'avait été obtenu jusqu'ici que pour le coefficient de corrélation.

Si les séries sont à liaisons internes, le problème est plus compliqué. Il faut alors faire intervenir, pour sa solution, des hypothèses sur la structure de ces liaisons.

Ces hypothèses doivent généraliser l'hypothèse d'indépendance faite en premier lieu.

On considérera des chaînes de variables aléatoires, liées par deux, par trois etc. Les liaisons dans la chaîne s'affaiblissent avec la distance.

Si l'on introduit l'hypothèse supplémentaire des régressions linéaires, on est amené, de la façon la plus naturelle à la décroissance géométrique des coefficients de corrélation, envisagée par EGON S. PEARSON et K. PEARSON, et à ses généralisations.

La solution générale du problème devient alors possible, par l'emploi des formules d'EGON S. PEARSON et ANDERSON.

Le lecteur remarquera que la raison de tous ces résultats est la propriété générale de la fonction caractéristique d'une loi de probabilité de fournir immédiatement la fonction caractéristique de combinaisons linéaires quelconques des variables.

Les différences d'ordre quelconque sont de telles combinaisons linéaires, mais la méthode, une fois comprise ainsi, s'applique à des cas beaucoup plus généraux.

V. ROMANOWSKY

On the moments of means of functions of one and more random variables

1. *Introductory.* — Let x be a random variable accepting the values

(1) $$x_1, x_2, \ldots, x_n$$

with the corresponding probabilities

(2) $$p_1, p_2, \ldots, p_n.$$

We shall suppose that these values are incompatible one with another and that x cannot accept any other values. Then $\sum_{i=1}^{n} p_i = 1$.

The values (1) of x and their probabilities (2) represent the law of distribution of x or, shorter, the distribution of x, which we shall denote with $D(x)$.

We shall consider further functions

$$\varphi_h(x) \quad (h = 1, 2, \ldots, q)$$

allways determined for the values (1) and one valued for them, so that we shall omit this indication, speaking simply of functions $\varphi_h(x)$.

Let now be given a sample, consisting of s observations which are mutually independent and which give the values (1) with the absolute frequencies

$$a_1, a_2, \ldots, a_n$$

$\left(\sum_{i=1}^{n} a_i = s\right)$. We shall denote this sample with S and suppose that the distribution $D(x)$ is unaltered during our observations.

The main aim of this paper consists in the research of various moments for quantities constructed with the values of functions $\varphi_h(x)$ for the samples analoguous to the sample S. Together with these moments we shall find some general theorem concerning the distributions of the quantities mentioned.

2. *Moments of the mean of one function.* — Let us consider at first one function $\varphi(x)$ and its mean $\frac{1}{s}\Sigma a_i \varphi(x_i)$ for the sample S. We shall denote this mean with $M\varphi$ or $\bar{\varphi}$.

(3) $$M\varphi = \bar{\varphi} = \frac{1}{s}\sum_{i=1}^{n} a_i \varphi(x_i) .$$

By $E\bar{\varphi}^h$ we shall denote the mathematical expectation of $\bar{\varphi}^h$ and shall name it the h^{th} moment of $\bar{\varphi}^h$. Thus

(4) $$E\bar{\varphi}^h = E\ (M\varphi)^h = \sum_{\bar{\varphi}} P_{\bar{\varphi}}\, \bar{\varphi}^h ,$$

denoting with $P_{\bar{\varphi}}$ the probability of any special value of $\bar{\varphi}$ for any special sample S and summing with regard to all possible values of $\bar{\varphi}$ in all possible samples S of the size s.

The most convenient manner to research the moments $E\bar{\varphi}^h$ is to construct their generating function which we shall denote with $G\ (E\bar{\varphi}^h)$.

In order to find $G\ (E\bar{\varphi}^h)$ let us consider the function

(5) $$U = \left(\sum_{i=1}^{n} p_i\, e^{\frac{t}{s}\varphi(x_i)}\right)^s .$$

Developping the right side we find

$$U(t) = \Sigma \frac{s!}{a_1!\ a_2! \ldots a_n!} p_1^{a_1} p_2^{a_2} \ldots p_n^{a_n} e^{t \cdot \frac{1}{s} \Sigma a_i \varphi(x_i)}$$

where the sum is to be taken with regard to all positive integer values of a_i such that

$$a_1 + a_2 + \ldots + a_n = s .$$

But, evidently,

$$\frac{s!}{a_1!\ a_2! \ldots a_n!} p_1^{a_1} p_2^{a_2} \ldots p_n^{a_n}$$

is the probability of the sample S and, therefore, of $\bar{\varphi}$. We see that

$$U(t) = \sum_{\bar{\varphi}} P_{\bar{\varphi}}\, e^{t\bar{\varphi}}$$

Now, differentiating this relation h times with regard to t and putting thereupon $t = o$, we find

(6) $$\left[\frac{d^h\, U(t)}{dt^h}\right]_{t=o} = \sum_{\bar{\varphi}} P_{\bar{\varphi}}\, \bar{\varphi}^h = e\,\bar{\varphi}^h$$

We see in this manner that (5) represents the generating function of the moments $E\bar{\varphi}^h$:

(7) $$G(E\bar{\varphi}^h) = \left(\sum_{i=1}^{n} p_i\, e^{\frac{t}{s}\varphi(x_i)}\right)^s.$$

Supposing that the elementary moments

(8) $$E\varphi^k = \sum_{i=1}^{n} p_i \varphi^k(x_i) \quad (k = 1, 2, 3, \ldots)$$

can be evaluated, we find easily with the aid of (6) and (7) the different moments $E\bar{\varphi}^h$ expressed in the quantities (8). Indeed, putting $u = \Sigma\, p_i\, e^{\frac{t}{s}\varphi(x_i)}$, we find:

$$E\bar{\varphi} = \left[\frac{dU}{dt}\right]_{t=0} = \left[s\, u^{s-1}\, \Sigma\, p_i \frac{\varphi(x_i)}{s}\, e^{\frac{t}{s}\varphi(x_i)}\right]_{t=0}$$

$$= s\, \Sigma\, p_i \frac{\varphi(x_i)}{s} = E\,\varphi;$$

$$E\bar{\varphi}^2 = \left[\frac{d^2 U}{dt^2}\right]_{t=0} = \Big[s\,(s-1)\, u^{s-2}\, (\Sigma\, p_i \frac{\varphi(x_i)}{s}\, e^{\frac{t}{s}\varphi(x_i)})^2$$

$$+ s\, u^{s-1}\, \Sigma\, p_i \left(\frac{\varphi(x_i)}{s}\right)^2 e^{\frac{t}{s}\varphi(x_i)}\Big]_{t=0}$$

$$= s\,(s-1)\,(\Sigma\, p_i \frac{\varphi(x_i)}{s})^2 + s\, \Sigma\, p_i \left(\frac{\varphi(x_i)}{s}\right)^2$$

$$= \frac{s-1}{s}(E\varphi)^2 + \frac{1}{s}\,\varepsilon\,\varphi^2;$$

$$E\bar{\varphi}^3 = \left[\frac{d^3 U}{dt^3}\right]_{t=0} = s\,(s-1)\,(s-2)\left(\Sigma\, p_i \frac{\varphi(x_i)}{s}\right)^3$$

$$+ 3\,s\,(s-1)\, \Sigma\, p_i \frac{\varphi(x_i)}{s}\, \Sigma\, p_i \left(\frac{\varphi(x_i)}{s}\right)^2 + s\, \Sigma\, p_i) \left(\frac{\varphi(x_i)}{s}\right)^3$$

$$= \frac{(s-1)\,(s-2)}{s^2}(E\varphi)^3 + \frac{3\,(s-1)}{s^2} E\varphi\, E\varphi^2 + \frac{1}{s^2} E\varphi^3;$$

$$E\bar{\varphi}^4 = \frac{(s-1)(s-2)(s-3)}{s^3}(E\varphi)^4 + \frac{6(s-1)(s-2)}{s^3}(E\varphi)^2 E\varphi^2$$

$$+ \frac{3(s-1)}{s^3}(E\varphi^2)^2 + \frac{4(s-1)}{s^3} E\varphi\, E\varphi^3 + \frac{1}{s^3} E\varphi^4$$

and so on.

3. *An algorithm for the research of the moments* $E\bar{\varphi}^k$. — Introducing the notations

$$(9) \qquad \alpha_k = s^{-k} E\varphi^k = \left[\frac{d^k u}{dt^k}\right]_{t=0} \qquad (k = 1, 2, 3, \ldots)$$

and

$$(10) \qquad s_k = s(s-1)(s-2)\ldots(s-k+1), \quad s_1 = s,$$

we can rewrite the expressions for $E\bar{\varphi}$, $E\bar{\varphi}^2$, $E\bar{\varphi}^3$ and $E\bar{\varphi}^4$, obtained in the previous paragraph, as follows:

$$E\bar{\varphi} = s_1\alpha_1, \quad E\bar{\varphi}^2 = s_2\alpha_1^2 + s_1\alpha_2,$$
$$E\bar{\varphi}^3 = s_3\alpha_1^3 + 3s_2\alpha_1\alpha_2 + s_1\alpha_3,$$
$$E\bar{\varphi}^4 = s_4\alpha_1^4 + 6s_3\alpha_1^2\alpha_2 + 3s_2\alpha_2^2 + 4s_2\alpha_1\alpha_3 + s_1\alpha_4.$$

It is not difficult to deduce from these expressions the law by means of which we can write $E\bar{\varphi}^{h+1}$ from $E\bar{\varphi}^h$. This law can be formulated very simply if we introduce an operation which is analogous to differentiation and which we shall denote with O. This operation is defined as follows:

I. *It satisfies the addition and multiplication theorems of the differentiation with regard to one variable,* i. e., for example,

$$O(u + v - w) = O(u) + O(v) - O(w),$$
$$O(u v w) = v w\, O(u) + w u\, O(v) + u v\, O(w),$$

u, v, w being some functions to which the operation O is applicable.

II. *It satisfies such special conditions:*

$$O(s_k) = s_{k+1}\alpha_1 \quad (k = 1, 2, 3, \ldots);$$
$$O(\alpha_k^m) = m\,\alpha_k^{m-1}\alpha_{k+1} \quad (k, m = 1, 2, 3, \ldots);$$
$$O(A) = 0 \textit{ if } A \textit{ is a numerical constant.}$$

These special conditions define the functions to which the operation O is applicable.

We can write now how $E\overline{\varphi}^{h+1}$ can be obtained from $E\overline{\varphi}^h$. It is, namely,

$$(11) \qquad E\overline{\varphi}^{h+1} = O\,(E\overline{\varphi}^h)\,.$$

Indeed, the derivative $\frac{d^h U}{dt^h}$ represents a sum of terms such as

$$(12) \quad A s_k\, u^{s-k} \left(s^{-1} \Sigma\, p_i\, (x_i)\; e^{\frac{t}{s}\varphi(x_i)}\right)^{h_1} \left(s^{-2} \Sigma p_i\, \varphi^2\,(x_i)\, e^{\frac{t}{s}\varphi(x_i)}\right)^{h_2} \ldots,$$

where A is some definite integer number. Putting here $t = o$ we see that $E\overline{\varphi}^h$ is the corresponding sum of the terms such as

$$(13) \qquad A s_k\, \alpha_1^{h_1}\, \alpha_2^{h_2} \ldots .$$

Now, in order to find $E\overline{\varphi}^{h+1}$, we must differentiate the sum of terms (12) with regard to t and put thereupon $t = o$. Thus we see that $E\overline{\varphi}^{h+1}$ is a sum of expressions

$$A s_{k+1}\, \alpha_1^{h_1+1}\, \alpha_2^{h_2} \ldots + A s_k\, h_1\, \alpha_1^{h_1-1}\, \alpha_2^{h_2+1} \ldots + \ldots,$$

which, evidently, is equal to $O\,(A s_k\, \alpha_1^{h_1}\, \alpha_2^{h_2} \ldots)$. We verify in this manner that (11) is true.

The first moment $E\overline{\varphi}$ is

$$E\overline{\varphi} = E\varphi = s_1\, \alpha_1$$

and all other moments $E\overline{\varphi}^h$ can be found one after another with the aid of (11). For example:

$$\begin{aligned}
E\overline{\varphi}^2 &= O\,(E\overline{\varphi}) = O\,(s_1\,\alpha_1) = \alpha_1\, O\,(s_1) + s_1\, O\,(\alpha_1)\\
&= s_2\,\alpha_1^2 + s_1\,\alpha_2\,;\\
E\overline{\varphi}^3 &= O\,(E\overline{\varphi}^2) = O\,(s_2\,\alpha_1^2 + s_1\,\alpha_2)\\
&= s_3\,\alpha_1^3 + 2\,s_2\,\alpha_1\,\alpha_2 + s_2\,\alpha_1\,\alpha_2 + s_1\,\alpha_3\\
&= s_3\,\alpha_1^3 + 3\,s_2\,\alpha_1\,\alpha_2 + s_1\,\alpha_3\,;\\
E\overline{\varphi}^4 &= O\,(s_3\,\alpha_1^3 + 3\,s_2\,\alpha_1\,\alpha_2 + s_1\,\alpha_3)\\
&= s_4\,\alpha_1^4 + 6\,s_3\,\alpha_1^2\,\alpha_2 + 3\,s_2\,\alpha_2^2 + 4\,s_2\,\alpha_1\,\alpha_3 + s_1\,\alpha_4\,.
\end{aligned}$$

We shall name the operation O *zero-differentiation.* Repeating this operation k times successively we shall have *zero-differentiation of order* k, which we shall denote with O^k. Thus we have

$$(14) \qquad O^k\,(u) = O\,[O^{k-1}\,(u)]\,,\; O^1\,(u) = O\,(u)\,.$$

It is clear that

$$(15) \qquad O^{k+m}\,(u) = O^k\,[O^m\,(u)] = O^m\,[O^k\,(u)]\,.$$

It is also easy to establish the following relation

$$O^k\ (uv) = \sum_{k_1=0}^{k} \frac{k!}{k_1!\ k_2!}\ O^{k_1}(u)\ O^{k_2}(v) \tag{16}$$

$$(k_1 + k_2 = k;\ O^\circ(u) = u\,,\ O^\circ(v) = v)$$

which is nothing else as the well known LEIBNITZ's theorem extended on the *zero-differentiation.*

We shall show now an application of the relation (16). We have

$$E\overline{\varphi}^{h+1} = O\ (E\overline{\varphi}^h) = O^2\ (E\overline{\varphi}^{h-1}) = \ldots = O^h\ (E\overline{\varphi})$$

or

$$E\overline{\varphi}^{h+1} = O^h\ (s_1\ \alpha_1)$$

whence, applying (16),

$$E\overline{\varphi}^{h+1} = \sum_{h_1=0}^{h} \frac{h!}{h_1!\ h_2!}\ O^{h_1}(s_1)\ O^{h_2}(\alpha_1) \quad (h_1 + h_2 = h)\ .$$

But

$$O^{h_2}(\alpha_1) = \alpha_{h_2+1}$$

and

$$\begin{aligned} O^{h_1}(s_1) &= O^{h_1-1}\ (s_2\ \alpha_1) \\ &= O^{h_1-2}\ (s_3\ \alpha_1^2 + s_2\ \alpha_2) \\ &= O^{h_1-3}\ (s_4\ \alpha_1^3 + 3\ s_3\ \alpha_1\ \alpha_2 + s_2\ \alpha_3) \\ &= \ldots \\ &= E'\overline{\varphi}^{h_1}\,, \end{aligned}$$

where $E'\overline{\varphi}^{h_1}$ is defined in the following way:
if we have

$$E\overline{\varphi}^{h_1} = \Sigma\ A\ s_k\ \alpha_1^{k_1}\ \alpha_2^{k_2} \ldots \varphi_r^{k_r}\,,$$

then

$$E'\overline{\varphi}^{h_1} = \Sigma\ A\ s_{k+1}\ \alpha_1^{k_1}\ \alpha_2^{k_2} \ldots \alpha_r^{k_r}\,,$$

i. e. we obtain $E'\overline{\varphi}^{h_1}$ from $E\overline{\varphi}^{h_1}$ by replacing in its terms every factorial s_k with s_{k+1}.

Thus we see that

$$E\overline{\varphi}^{h+1} = \sum_{h_1=0}^{h} \frac{h!}{h_1!\ h_2!}\ \alpha_{h_2+1}\ E'\overline{\varphi}^{h_1}.\ \ (E'\overline{\varphi}^\circ = s_1) \tag{17}$$

For example:

$$\begin{aligned} E\overline{\varphi}^4 &= \alpha_1\ E'\overline{\varphi}^3 + 3\,\alpha_2\ E'\overline{\varphi}^2 + 3\,\alpha_3\ E'\overline{\varphi} + s_1\ \alpha_4 \\ &= \alpha_1\ (s_4\ \alpha_1^3 + 3\ s_3\ \alpha_1\ \alpha_2 + s_2\ \alpha_3) \\ &+ 3\,\alpha_2\ (s_3\ \alpha_1^2 + s_2\ \alpha_2) + 3\,\alpha_3\,.\,s_2\ \alpha_1 + s_1\ \alpha_4 \\ &= s^4\ \alpha_1^4 + 6\ s_3\ \alpha_1^2\ \alpha_2 + 3\ s_2\ \alpha_2^2 + 4\ s_2\ \alpha_1\ \alpha_3 + s_1\ \alpha_4\,, \end{aligned}$$

as it must be.

4. *A general independent expression for $E\bar{\varphi}^h$.* — We shall find now a general expression for $E\bar{\varphi}^h$, which is independent of $E\bar{\varphi}, E\bar{\varphi}^2, \ldots, E\bar{\varphi}^{h-1}$ and which is very important for our further investigation.

To find it we apply the following generalisation of the LEIBNITZ's theorem which can be established without pain and which we shall not demonstrate:

If $u_1, u_2, \ldots, u_s$ *be some functions of a variable* t, *having all derivatives of orders* $1, 2, \ldots, h$, *then*

$$(18) \quad \frac{d^h (u_1 u_2 \ldots u_s)}{dt^h} = \Sigma \frac{h!}{h_1!\, h_2! \ldots h_s!} \frac{d^{h_1} u_1}{dt^{h_1}} \frac{d^{h_2} u_2}{dt^{h_2}} \cdots \frac{d^{h_s} u_s}{dt^{h_s}},$$

where the sum is to be taken with regard to all positive integer values (zero including) of $h_1, h_2, \ldots, h_s$ *such that* $h_1 + h_2 + \ldots h_s = h$.

Put in this formula

$$u_1 = u_2 = \ldots = u_s = u = \Sigma p_i e^{\frac{t}{s}\varphi(x_i)},$$

then

$$E\bar{\varphi}^h = \left[\frac{d_h u^s}{dt^h}\right]_{t=0} = \Sigma \frac{h!}{h_1!\, h_2! \ldots h_s!} \left[\frac{d^{h_1} u}{dt^{h_1}} \frac{d^{h_2} u}{dt^{h_2}} \cdots \frac{d^{h_s} u}{dt^{h_s}}\right]_{t=0},$$

or

$$(19) \qquad E\bar{\varphi}^h = \Sigma \frac{h!}{h_1!\, h_2! \ldots h_s!} \alpha_{h_1} \alpha_{h_2} \ldots \alpha_{h_s},$$

for

$$\left[\frac{d^{h_i} u}{dt^{h_i}}\right]_{t=0} = \alpha_{h_i},$$

as it follows from (9).

Remarque now that in the sum of (19) various h_i can accept the identical values and that therefore we shall have identical terms in this sum. Let it been among the numbers h_i, satisfying the condition $\Sigma h_i = h$, ν_1 having the value λ_1, ν_2 having the value $\lambda_2, \ldots,$ ν_r having the value λ_r, $r \leq s$. Then the term

$$\frac{h!}{h_1!\, h_2! \ldots h_s!} \alpha_{h_1} \alpha_{h_2} \ldots \alpha_{h_s} = \frac{h!}{(\lambda_1!)^{\nu_r} \ldots (\lambda_r!)^{\nu_r}} \alpha_{\lambda_1}^{\nu_1} \ldots \alpha_{\lambda_r}^{\nu_r}$$

is repeated in the sum of (19)

$$\frac{s(s-1)\ldots(s-\nu+1)}{\nu_1!\, \nu_2! \ldots \nu\,!} = \frac{s_\nu}{\nu_1!\, \nu_2! \ldots \nu_r!}$$

$$(\nu = \nu_1 + \nu_2 + \ldots + \nu_r)$$

times. In this way we obtain the following general expression for $E\bar{\varphi}^h$:

(20) $$E\bar{\varphi}^h = \Sigma \frac{h!}{(\lambda_1!)^{\nu_1}\dots(\lambda_r!)^{\nu_r}} \frac{s_\nu}{\nu_1!\dots\nu_r!} \alpha_{\lambda_1}{}^{\nu_1}\dots\alpha_{\lambda_r}{}^{\nu_r},$$

where

(21) $$\alpha_\lambda = s^{-\lambda} E\varphi^\lambda;\ \nu = \nu_1 + \nu_2 + \dots + \nu_r$$

and the sum is to be taken for all positive integers ν_i, λ_i such that

(22) $$\nu_1\lambda_1 + \nu_2\lambda_2 + \dots + \nu_r\lambda_r = h \ (r \leq s),$$

whereas $\lambda_1, \lambda_2, \dots, \lambda_r$ must be all different. This expression can be written also in the form:

(23) $$E\bar{\varphi}^h = s^{-h}\Sigma \frac{h!}{(\lambda_1!)^{\nu_1}\dots(\lambda_r!)^{\nu_r}} \frac{s_\nu}{\nu_1!\dots\nu_r!} (E\varphi^{\lambda_1})^{\nu_1}\dots(E\varphi^{\lambda_r})^{\nu_r}.$$

Example. We apply this formula for the research of $E\bar{\varphi}^4$. For this aim we find the following table of the required positive integer solutions of the equation (22) together with the values of number ν:

λ_1	λ_2	ν_1	ν_2	ν
1	0	4	0	4
1	2	2	1	3
2	0	2	0	2
3	1	1	1	2
4	0	1	1	1

With the aid of this table and of (20) we have now:

$$E\bar{\varphi}^4 = \frac{4!}{(1!)^4}\frac{s_4}{4!}\alpha_1^4 + \frac{4!}{(1!)^2\,2!}\frac{s_3}{2!\,1!}\alpha_1^2\alpha_2 + \frac{4!}{(2!)^2}\frac{s_2}{2!}\alpha_2^2 +$$

$$+ \frac{4!}{3!\,1!}\frac{s_2}{1!\,1!}\alpha_3\alpha_1 + \frac{4!}{4!}\frac{s_1}{1!}\alpha_4$$

$$= s_4\alpha_1^4 + 6\,s_3\,\alpha_1^2\alpha_2 + 3\,s_2\,\alpha_2^2 + 4\,s_2\,\alpha_1\alpha_3 + s_1\alpha_4.$$

Let us make some remarks concerning the general formula (20).

1st *remark.* There are distributions $D(x)$ for which

$$\alpha_1 = s^{-1}E\varphi = 0.$$

Then, evidently, we must take for the solutions of (22) the values of $\lambda_1, \lambda_2, \ldots, \lambda_r$ different from 2. In this case $E\bar{\varphi}^h$ is dependent only on $E\varphi^2, E\varphi^3, \ldots, E\varphi^4$.

2nd remark. If the distribution $D(x)$ is such that

$$E\varphi = E\varphi^3 = E\bar{\varphi}^5 = \ldots = E\varphi^{2k+1} = 0,$$

then we shall have also

$$E\bar{\varphi} = E\bar{\varphi}^3 = E\bar{\varphi}^5 = \ldots = E\bar{\varphi}^{2k+1} = 0,$$

as it is easy to prove.

3rd remark. If all odd moments of φ are zero, then all odd moments of $\bar{\varphi}$ are also zero.

5. *The first eight moments of* $\bar{\varphi}$. — With the aid of (20) or (11, we find the following expressions for the first eight moments of $\bar{\varphi}$:

$$E\bar{\varphi} = s_1 \alpha_1,$$
$$E\bar{\varphi}^2 = s_2 \alpha_1^2 + s_1 \alpha_2,$$
$$E\bar{\varphi}^3 = s_3 \alpha_1^3 + 3 s_2 \alpha_1 \alpha_3 + s_1 \alpha_3,$$
$$E\bar{\varphi}^4 = s_4 \alpha_1^4 + 6 s_3 \alpha_1^2 \alpha_2 + s_2 (3\alpha_2^2 + 4\alpha_1 \alpha_3) + s_1 \alpha_4,$$
$$E\bar{\varphi}^5 = s_5 \alpha_1^5 + 10 s_4 \alpha_1^3 \alpha_2 + s_3 (15 \alpha_1 \alpha_2^2 + 10 \alpha_1^2 \alpha_3) + s_2 (10 \alpha_2 \alpha_3 + 5 \alpha_1 \alpha_4) + s_1 \alpha_5,$$
$$E\bar{\varphi}^6 = s_6 \alpha_1^6 + 15 s_5 \alpha_1^4 \alpha_2 + s_4 (45 \alpha_1^2 \alpha_2^2 + 20 \alpha_1^3 \alpha_3) + s_3 (15 \alpha_2^3 + 60 \alpha_1 \alpha_2 \alpha_3 + 15 \alpha_1^2 \alpha_4) + s_2 (10 \alpha_3^2 + 15 \alpha_2 \alpha_4 + 6 \alpha_1 \alpha_5) + s_1 \alpha_6,$$
$$E\bar{\varphi}^7 = s_7 \alpha_1^7 + 21 s_6 \alpha_1^5 \alpha_2 + s_5 (105 \alpha_1^3 \alpha_2^2 + 35 \alpha_1^4 \alpha_3) + s_4 (105 \alpha_1 \alpha_2^3 + 210 \alpha_1^2 \alpha_2 \alpha_3 + 35 \alpha_1^3 \alpha_4) + s_3 (105 \alpha_2^2 \alpha_3 + 70 \alpha_1 \alpha_3^2 + 105 \alpha_1 \alpha_2 \alpha_4 + 21 \alpha_1^2 \alpha_5) + s_2 (35 \alpha_3 \alpha_4 + 21 \alpha_2 \alpha_5 + 7 \alpha_1 \alpha_6) + s_1 \alpha_7,$$
$$E\bar{\varphi}^8 = s_8 \alpha_1^8 + 28 s_7 \alpha_1^6 \alpha_2 + s_6 (210 \alpha_1^4 \alpha_2^2 + 56 \alpha_1^5 \alpha_3) + s_5 (420 \alpha_1^2 \alpha_2^3 + 560 \alpha_1^3 \alpha_2 \alpha_3 + 70 \alpha_1^4 \alpha_4) + s_4 (105\alpha_2^4 + 840\alpha_1 \alpha_2^2 \alpha_3 + 280\alpha_1^2 \alpha_3^2 + 420\alpha_1^2 \alpha_2 \alpha_4 + 56\alpha_1^3 \alpha_5) + s_3 (280\alpha_2 \alpha_3^2 + 210\alpha_2^2 \alpha_4 + 280\alpha_1 \alpha_3 \alpha_4 + 168\alpha_1 \alpha_2^2 \alpha_3 + 28\alpha_1^2 \alpha_6) + s_2 (35 \alpha_4^2 + 56 \alpha_3 \alpha_5 + 28 \alpha_2 \alpha_6 + 8 \alpha_1 \alpha_7) + s_1 \alpha_8.$$

If $E\varphi = 0$, we shall have from these formulae:

$$E\bar{\varphi} = 0, \quad E\bar{\varphi}^2 = s_1 \alpha_2, \quad E\bar{\varphi}^3 = s_1 \alpha_3,$$
$$E\bar{\varphi}^4 = 3 s_2 \alpha_2^2 + s_1 \alpha_4,$$
$$E\bar{\varphi}^5 = 10 s_2 \alpha_2 \alpha_3 + s_1 \alpha_5,$$
$$E\bar{\varphi}^6 = 15 s_3 \alpha_2^3 + s_2 (10 \alpha_3^2 + 15 \alpha_2 \alpha_4) + s_1 \alpha_6,$$
$$E\bar{\varphi}^7 = 105 s_3 \alpha_2^2 \alpha_3 + s_2 (35 \alpha_3 \alpha_4 + 21 \alpha_2 \alpha_5) + s_1 \alpha_7,$$
$$E\bar{\varphi}_8 = 105 s_4 \alpha_2^4 + s_3 (280 \alpha_2 \alpha_3^2 + 210 \alpha_2^2 \alpha_4) + s_2 (35 \alpha_4^2 + 56 \alpha_3 \alpha_5 + 28 \alpha_2 \alpha_6) + s_1 \alpha_8.$$

6. *The distribution of $\overline{\varphi}$ for the case $E\varphi = 0$.* — We shall now demonstrate the following fondamental theorem:

If $E\varphi = 0$, the probability of the inequalities

$$t_2\sqrt{2\,E\overline{\varphi}^2} < \overline{\varphi} < t_2\sqrt{2\,E\overline{\varphi}^1}$$

tends for $s \to \infty$ to the limit

$$\frac{1}{\sqrt{\pi}} \int_{t_1}^{t_2} e^{-t^2}\, dt\,,$$

whatever real constant numbers t_1 and $t_2 > t_1$ may be.

Using the TCHEBYSHEFF-MARKOFF's theorem we can very easily prove our theorem. The TCHEBYSHEFF-MARKOFF's theorem can be stated as follows:

If a random variable x *whose distribution is dependent on the positive integer* s *is such that*

$$(24) \qquad \lim_{s \to \infty} Ex^k = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^k e^{-t^2}\, dt \quad \text{for} \quad k = 0, 1, 2, 3, \ldots,$$

then the probability of the inequalities

$$t_1 < x < t_2$$

tends for $s \to \infty$ to the limit

$$\frac{1}{\sqrt{\pi}} \int_{t_1}^{t_2} e^{-t^2}\, dt\,,$$

whatever real constants t_1 *and* $t_2 > t_1$ *may be.*

Now, when $E\varphi = 0$, it is easy to show that we shall have

$$\lim_{s \to \infty} E\left(\frac{\overline{\varphi}}{\sqrt{2\,E\overline{\varphi}^2}}\right)^k = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^k e^{-t^2}\, dt \ \text{for} \ k = 0, 1, 2, 3, \ldots,$$

wherefrom it follows at once the exactness of our fondamental theorem.

Indeed, we conclude from (20) that the highest in s terms of $E\bar{\varphi}^{2h}$ and $E\bar{\varphi}^{2h+1}$ are correspondingly

$$\frac{(2h)!}{2^h}\frac{s_h}{h!}\alpha_2{}^h = \frac{(2h)!}{2^h\,h!}\frac{s_h}{s^{2h}}(E\varphi^2)^h$$

and

$$\frac{(2h+1)!}{2^{h-1}.3!}\frac{s_h}{(h-1)!}\alpha_2{}^{h-1}\alpha_3 = \frac{(2h+1)!}{2^{h-1}.3!\,(h-1)!}\frac{s_h}{s^{2h+1}}(E\varphi^2)^{h-1}.E\varphi^3$$

Therefore

$$\lim_{s\to\infty} E\left(\frac{\bar{\varphi}}{\sqrt{2\,E\bar{\varphi}^2}}\right)^{2h} = \lim_{s\to\infty}\frac{E\bar{\varphi}^{2h}}{(2\,E\bar{\varphi}^2)^h}$$

$$= \lim_{s\to\infty}\frac{(2\,h)!}{2^{2h}\,h!}\frac{s_h\,s^h}{s^{2h}}$$

$$= \frac{(2\,h)!}{2^{24}\,h!} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} t^{2h}\,e^{-t^2}\,dt;$$

$$\lim_{s\to\infty} E\left(\frac{\bar{\varphi}}{\sqrt{2\,E\bar{\varphi}^2}}\right)^{2h+1} = \lim_{s\to\infty}\frac{E\bar{\varphi}^{2h+1}}{(2\,E\bar{\varphi}^2)^{h+1/2}}$$

$$= \lim_{s\to\infty}\frac{(2\,h+1)!}{2^{2h-1/2}.\,3!\,(h-1)!}\frac{s_h\,s^{h+1/2}}{s^{2h+1}}\frac{E\varphi^3}{(E\varphi^2)^{3/2}}$$

$$= 0 = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} t^{2h+1}\,e^{-t^2}\,dt$$

for $h = 0, 1, 2, 3, \ldots$. Accordingly (24) is proved and our theorem is exact.

Denoting with $D(\bar{\varphi})$ the distribution law of $\bar{\varphi}$ we deduce from the theorem just proved that for s sufficiently great we can write approximately

(25) $$D(\bar{\varphi}) = \frac{1}{\sqrt{2\pi\,E\bar{\varphi}^2}}\,e^{-\bar{\varphi}^2/2E\bar{\varphi}^2}$$

$$= \sqrt{\frac{s}{2\pi\,E\varphi^2}}\,e^{-s\bar{\varphi}/2E\varphi^2},$$

if $E\varphi = 0$.

Let now be

$$\varphi_0 = E\varphi \neq 0.$$

Then instead of φ we can consider the function $\omega(x) = \varphi(x) - \varphi_0$ for which we shall have

$$E\omega = E\varphi - \varphi_0 = 0.$$

Accordingly, our theorem will be applicable to this function and we shall have approximately

$$(26) \qquad D(\bar{\varphi}) = \sqrt{\frac{s}{2\pi\,(E\bar{\varphi}^2 - \varphi_0^2)}}\; e^{-s(\bar{\varphi}-\varphi_0)^2/2\,(E\varphi^2-\varphi_0^2)},$$

when s is great.

7. *Some special cases of the function* $\varphi(x)$. — Giving to $\varphi(x)$ special form we can receive great variety of cases which have important statistical applications. We shall consider in this paragraph some of such special forms of $\varphi(x)$.

a) $\varphi(x) = x$. In this case, assuming the notations of AL. A. TCHOUPROFF, we shall have

$$E\varphi^k = Ex^k = m_k;$$
$$\alpha_k = s^{-k}\, m_k,\ \bar{\varphi} = Mx = \bar{x}$$
$$E\bar{\varphi}^k = E\bar{x}^k = m_{k\,(s)}$$

and our fondamental formulae for the calculation of $E\bar{\varphi}^h$, (11), (17) and (20), give us for the present case:

$$m_{h+1\,(s)} = O\,(m_{h\,(s)});$$

$$m_{h+1\,(s)} = \alpha_1\, m'_{h\,(s)} + \frac{h}{1}\,\alpha_2\, m'_{h-1\,(s)} + \ldots + S_1\,\alpha_{h+1}$$

$$= s^{-1}\, m_1\, m'_{h\,(s)} + \frac{h}{1}\, s^{-2}\, m_2\, m'_{h-1\,(s)} + \ldots + s^{-h}\, m_{h+1};$$

$$m_{h\,(s)} = s^{-h}\,\Sigma\, \frac{h!}{(\lambda_1!)^{\nu_1}\ldots(\lambda_r!)^{\nu_r}}\; \frac{s_\nu}{\nu_1!\ldots\nu_r!}\; m_{\lambda_1}^{\ \nu_1}\ldots m_{\lambda_r}^{\ \nu_r}.$$

The fondamental theorem gives in this case

$$D(\bar{x}) \backsim \sqrt{\frac{s}{2\pi\mu_2}}\; e^{-s(\bar{x}-m_1)^2/2\mu_2} \text{ for } s \to \infty$$
$$(\mu_2 = m_2 - m_1^2).$$

b) $\varphi(x) = x - m_1$. Then

$$E\varphi^k = E\,(x-m_1)^k = \mu_k;\ \mu_1 = 0;\ \alpha_k = s^{-k}\,\mu_k;$$
$$\bar{\varphi} = \bar{x} - m_1;\ E\bar{\varphi}^h = E\,(\bar{x}-m_1)^h = \mu_{h\,(s)}$$

and it is easy to write now the fundamental formulae for $\mu_{h\,(s)}$.

We remark only the general formula

$$\mu_{h\,(s)} = s^{-h} \sum \frac{h!}{(\lambda_1 !)^{\nu_1} \dots (\lambda_r !)^{\nu_r}} \frac{s_\nu}{\nu_1 ! \dots \nu_r !} \mu_{\lambda_1}{}^{\nu_1} \dots \mu_{\lambda_r}{}^{\nu_r}$$

where, as now $\mu_1 = 0$, $\lambda_1, \lambda_2, \dots, \lambda_r$ take different values 2, 3, 4, ... satisfying the usual condition $\nu_1 \lambda_1 + \dots + \nu_r \lambda_r = h$.

c) $\varphi(x) = x^a$, a being any real number. In the case $a < 0$ the values $x_1, x_2, \dots, x_n$ must not include zero. We have now

$$E\varphi^k = Ex^{ka} = m_{ka}, \quad \alpha_k = s^{-k} m_{ka};$$

$$\bar{\varphi} = Mx^a, \quad E\bar{\varphi}^k = E\,[Mx^a].$$

d) $\varphi(x) = (x - m_1)^a$. Then

$$E\varphi^k = \mu_{ka}, \quad \alpha_k = s^{-k} \mu_{ka};$$

$$\bar{\varphi} = M\ (x - m_1)^a = \mu^1{}_{a\,(s)}$$

$$E\bar{\varphi}^k = E\mu^1{}_{a\,(s)}.$$

e) $\varphi(x) = (x - m_1)^a - \mu_a$. Then

$$E\varphi^k = E\,[(x - m_1)^a - \mu_a]^k = \sum_{g=0}^{k} (-1)^{k-g} \frac{k!}{g!\,(k-g)!} \mu_{ga}\, \mu_a{}^{k-g},$$

$$\alpha_k = s^{-k} E\varphi^k,$$

$$\bar{\varphi} = M\ [(x - m_1)^a - \mu^a] = \mu'_{a\,(s)} - \mu_a,$$

$$E\bar{\varphi}^k = E\ (\mu'_{a\,(s)} - \mu_a)^k.$$

For all these cases it is very easy to write down the general expression of $E\bar{\varphi}^k$ introducing in (20) the corresponding special values of α_k. Paragraph 5 gives for the cases considered the eight first moments. In all cases, for s sufficiently great, we can assume that the distribution of the mean of the corresponding special form of $\varphi(x)$ is approximately normal.

8. *Generalisations.* — We have considered the distribution $D(x)$ in which x assumes a finite number of different values with some determined probabilities. We shall now remove this restriction.

Let the values of x constitute some infinite set E of numbers, which may be enumerable or not. Let a and b be the lower and upper limits of these values, whereas the infinite values for a and b are not excluded. The distribution of x, $D(x)$, may be represented now by a function $f(x)$ with the following properties:

1. *$f(x)$ is defined and one valued in the interval* (a, b).
2. *$f(x)$ is continuous on the right at each value of* x *in* (a, b).
3. *$f(x)$ is not decreasing in* (a, b).
4. *$f(a) = 0$ and $f(b) = 1$.*

5. *The probability of the inequalities* $a \leq x \leq t$, *where* $a < t \leq b$, *is equal to the* STIELTJES's *integral* $\int_a^t df(x)$.

We shall consider now a function $\varphi(x)$, defined and one valued in (a, b) and suppose that do exist the integrals

$$\int_a^b [\varphi(x)]^k \, df(x) \quad (k = 1, 2, \ldots, N), \tag{26}$$

N being some definite integer number, and that the integral

$$u = \int_a^b e^{\frac{t}{s}\varphi(x)} \, df(x)$$

is existing also for some finite interval of the values of t including $t = 0$.

It is clear that the integrals (26) represent the moments $E\varphi^k(x)$ of the function $\varphi(x)$.

Let now be given some sample consisting of s independent observations of x giving

$$x_1, \; x_2, \; \ldots, \; x_s$$

and let it be

$$\overline{\varphi} = \frac{1}{s} \sum_{i=1}^{s} \varphi(x_i) = M\varphi(x)$$

Then it is easy to see that the generating function of the moments $E\overline{\varphi}^h$ $(h = 0, 1, 2, \ldots, N)$ is given by the relation

$$G(E\overline{\varphi}^h) = u^s = \left[\int_a^b e^{\frac{t}{s}\varphi(x)} \, df(x)\right]^s, \tag{27}$$

so that

$$E\overline{\varphi}^h = \left[\frac{d^h u^s}{dt^h}\right]_{t=0}.$$

It is also evident that we shall have the relations

$$E\overline{\varphi}^{h+1} = O(E\overline{\varphi}^h),$$

$$E\overline{\varphi}^{h+1} = \sum_{g=0}^{h} \frac{h!}{g!\,(h-g)!} \alpha_{h-g+1} \, E'\overline{\varphi}^h,$$

$$E\overline{\varphi}^h = \sum \frac{h!}{(\lambda_1 !)^{\nu_1} \ldots (\lambda !)^{\nu_r}} \frac{s_r}{\nu_1 ! \ldots \nu_r !} \alpha_{\lambda_1}^{\nu_1} \ldots \alpha_{\lambda_r}^{\nu_r},$$

which are identical with the relations (11), (17) and (20) and in which

$$\alpha_k = s^{-k} \int_a^b [\varphi(x)]^k \, df(x)$$

and O is the operation defined in § 3.

When the integrals (26) exist for all possible integer values of k and have finite values for all finite values of k then it is not difficult to see that the fundamental theorem of § 6 subsists also for the case now considered and that, for s sufficiently great, we can again write approximately

$$D(\overline{\varphi}) = \sqrt{\frac{s}{2\pi(E\varphi^2 - \varphi_0^2)}} \, e^{-s(\overline{\varphi} - \varphi_0)^2/2(E\varphi^2 - \varphi_0^2)},$$

where now

$$\varphi_0 = \int_a^b \varphi(x) \, df(x), \quad E\varphi^2 = \int_a^b [\varphi(x)]^2 \, df(x).$$

The fundamental theorem of § 6 was demonstrated by means of the TCHEBYSHEFF-MARKOFF's theorem. In the case of discrete distribution $D(x)$ considered in the previous paragraphs it is the most natural way because there exist in that case all moments $E\overline{\varphi}^h$ ($h = 0, 1, 2, 3, \dots$) and these moments have the properties required by the TCHEBYSCHEFF-MARKOFF's theorem. But in the present case of the distribution $D(x)$ of the most general kind the moments $E\overline{\varphi}^h$ do not exist allways and it is of great interest to obtain the distribution theorem for $D(\overline{\varphi})$ in possibly least restrictive conditions. We can do it with the aid of another celebrated theorem of the calculus of probabilities, that of LIAPOUNOFF.

The LIAPOUNOFF's theorem (*Nouvelle forme du théorème sur la limite de probabilité*, « Mémoires de l'Académie des Sciences de St. Pétersbourg », Classe physico-math., vol. XII, N. 5) can be formulated as follows:

« LIAPOUNOFF's » *theorem. Let* $u_1, u_2, u_3, \dots$ *be infinite series of independent random variables admitting the moments*

Eu_h, $E(u_h - Eu_h)^2$ and $E\,|u_h - Eu_h|^{2+\delta}$ ($h = 1, 2, 3, \dots$) (δ *being some positive real number*) *such that*

$$\frac{\sum_{h=1}^{s} E\,|u_h - Eu_h|^{2+\delta}}{\left[\sum_{h=1}^{s} E(u_h - Eu_h)^2\right]^{1+\frac{\delta}{2}}} \to 0 \text{ for } s \to \infty$$

Then the probability of inequalities

$$z_1 < \frac{\sum\limits_{h=1}^{s} (u_h - Eu_h)}{\sqrt{2 \sum\limits_{h=1}^{s} (u_h - Eu_h)^2}} < z_2$$

tends for $s \rightarrow \infty$ *to the limit*

$$\frac{1}{\sqrt{\pi}} \int\limits_{z_1}^{z_2} e^{-z^2}\, dz$$

whatever real constants z_1 *and* z_2 *may be.*

Now, basing on this theorem, it is easy to prove that if our distribution $D(x)$ and our function $\varphi(x)$, which we are considering in this paragraph, are such that are existing and are finite, the moments

$$\varphi_0 = E\varphi = \int\limits_a^b \varphi(x)\, df(x),$$

$$E(\varphi - \varphi_0)^2 = \int\limits_a^b [\varphi(x) - \varphi_0]^2\, df(x),$$

$$E\,|\varphi - \varphi_9|^{2+\delta} = \int\limits_a^b |\varphi(x) - \varphi_0|^{2+\delta}\, df(x),$$

δ being some positive real number, then the distribution $D(\bar{\varphi})$ of $\bar{\varphi}$ tends for $s \rightarrow \infty$ to be normal.

Indeed, putting

$$u_h = \varphi(x_h) \quad (h = 1, 2, 3, \ldots, s),$$

where x_h $(h = 1, 2, 3, \ldots, s)$ are the values of x which can be observed in any random sample of the size s, weobtain the random variables satisfying to all conditions of the LIAPOUNOFF's theorem: their number increases without limit with $s \rightarrow \infty$, they are independent and have the required moments with the required property, for we have

$$Eu_h = E\varphi(x_h) = E\varphi(x) = \varphi_0\,,$$

$$E\,(u_h - Eu_h)^2 = E\,[\varphi(x_h) - \varphi_0]^2 = E\,[\varphi(x) - \varphi_0]^2\,,$$

$$E\,|u^h - Eu_h|^{2+\delta} = E\,|\varphi(x_h) - \varphi_0|^{2+\delta} = E\,|\varphi(x) - \varphi_0|^{2+\delta}$$

$(h = 1, 2, 3, \ldots, s)$ and

$$\lim_{s \to \infty} \frac{\sum_{h=1}^{s} E\,|\,u_h - Eu_h\,|^{2+\delta}}{\left[\sum_{h=1}^{s} E\,(u_h - Eu_h)^2\right]^{1+\frac{\delta}{2}}} =$$

$$= \lim_{s \to \infty} \frac{E\,|\,\varphi(x) - \varphi_0\,|^{2+\delta}}{(E\,[\varphi(x) - \varphi\,]^2)^{1+\frac{\delta}{2}}} \; \frac{1}{s^{\delta/2}} = 0.$$

We see that the LIAPOUNOFF's theorem is applicable to our case when are existing the finite moments

$$\varphi_0, \; E\,(\varphi - \varphi_0)^2 \text{ and } E\,|\,\varphi - \varphi_0\,|^{2+\delta}$$

for some positive δ and that we can then write:

$$D\,(\bar{\varphi}) \backsim \sqrt{\frac{s}{2\pi E\,(\varphi - \varphi_0)^2}} \; e^{-s\,(\varphi - \varphi_0)^2/2\,E\,(\varphi - \varphi_0)^2}$$

for $s \to \infty$.

9. *The case of two and more functions of one random variable.* — Let us return to the suppositions of § 1 concerning the random variable x and consider now two functions $\varphi(x)$ and $\psi(x)$. Let their means for some sample S be

$$\bar{\varphi} = \frac{1}{s} \sum_{i=1}^{n} a_i\, \varphi(x_i), \quad \bar{\psi} = \frac{1}{s} \sum_{i=1}^{n} a_i\, \psi(x_i)$$

and let us take the product-moments of these means of the form

$$E\bar{\varphi}^h\, \bar{\psi}^g = \sum_{\bar{\varphi}, \bar{\psi}} P_{\bar{\varphi}, \bar{\psi}}\, \bar{\varphi}^h\, \bar{\psi}^g, \tag{28}$$

where $P_{\bar{\varphi}, \bar{\psi}}$ is the probability of $\bar{\varphi}$ and $\bar{\psi}$ and the sum must be taken for all possible values of these means.

It is almost evident that the generating function of the moments (28) is

$$G\,(E\bar{\varphi}^h\, \bar{\psi}^g) = u^s \equiv \left(\sum_{i=1}^{n} p_i\, e^{\frac{t}{s}\varphi(x_i) + \frac{\theta}{s}\psi(x_i)}\right)^s \tag{29}$$

so that

$$E\bar{\varphi}^h\, \bar{\psi}^g = \left[\frac{\delta^{h+g}\, u^s}{\delta t^h\, \delta\theta^g}\right]_{t=\theta=0} \tag{30}$$

Evaluating $E\bar{\varphi}^h \bar{\psi}^g$ for some initial values of h and g we come very easily to the conclusion that the moments $E\bar{\varphi}^h \bar{\psi}^g$ can be found very commodely by means of the recurrent relations

(31) $$E\bar{\varphi}^h \bar{\psi}^g = O^{10} \left(E\bar{\varphi}^{h-1}\bar{\psi}^g\right)$$

or

(31-*bis*) $$E\bar{\varphi}^h \bar{\psi}^g = O^{01} \left(E\bar{\varphi}^h \bar{\psi}^{g-1}\right)$$

where O^{10} and O^{01} are the signs of zero-differentiation in respect to two variables and represent operations satisfying the ordinary addition and multiplication rules for the differentiation of functions of two variables and the following special conditions:

(32) $$\begin{cases} O^{10}(s_k) = s_{k+1}\,\alpha_{10}\,, \quad O^{01}(s_k) = s_{k+1}\,\alpha_{01}\,; \\ O^{10}(\alpha_{kl}{}^m) = m\,\alpha_{kl}{}^{m-1}\,\alpha_{k+1,\,l}\,, \\ O^{01}(\alpha_{kl}{}^m) = m\,\alpha_{kl}{}^{m-1}\,\alpha_{k,\,l+1}\,, \\ O^{01}(A) = O^{01}(A) = 0\,, \end{cases}$$

where k, l, m are any positive integers,

$$s_k = s\,(s-1)\,\ldots\,(s-k+1)\,,$$

(33) $$\alpha_{kl} = s^{-k-l}\,E\varphi^k(x)\,\psi^l(x) = s^{-k-l}\sum_{i=1}^{n} p_i\,\varphi^k(x_i)\,\psi^l(x_i)$$

and A is any numerical constant.

Let us give some illustrations. By means of (30) we find at first

$$E\bar{\varphi} = s_1\,\alpha_{10}\,, \quad E\bar{\psi} = s_1\,\alpha_{01}$$

and further, by means of (31), (31-*bis*) and (32):

$$\begin{aligned} E\bar{\varphi}\,\bar{\psi} &= O^{10}(s_1\,\alpha_{01}) = O^{01}(s\,\alpha_{10}) = s_2\,\alpha_{10}\,\alpha_{01} + s_1\,\alpha_{11}\,; \\ E\bar{\varphi}^2\,\bar{\psi} &= O^{10}(E\bar{\varphi}\bar{\psi}) = O^{10}(s_2\,\alpha_{10}\,\alpha_{01} + s_1\,\alpha_{11}) \\ &= s_3\,\alpha_{10}{}^2\,\alpha_{01} + s_2\,\alpha_{20}\,\alpha_{01} + 2\,s_2\,\alpha_{10}\,\alpha_{11} + s_1\,\alpha_{21}\,; \\ E\bar{\varphi}\,\bar{\psi}^2 &= O^{01}(s_2\,\alpha_{10}\,\alpha_{01} + s_1\,\alpha_{11}) \\ &= s_3\,\alpha_{10}\,\alpha_{01}{}^2 + s_2\,\alpha_{10}\,\alpha_{02} + 2\,s_2\,\alpha_{01}\,\alpha_{11} + s_1\,\alpha_{12}\,; \\ E\bar{\varphi}^2\,\bar{\psi}^2 &= O^{01}(E\varphi^2\,\bar{\psi}) \\ &= O^{01}(s_3\,\alpha_{10}{}^2\,\alpha_{01} + s_2\,\alpha_{20}\,\alpha_{01} + 2\,s_2\,\alpha_{10}\,\alpha_{11} + s_1\,\alpha_{21}) \\ &= s_4\,\alpha_{10}{}^2\,\alpha_{01}{}^2 + s_3\,(\alpha_{10}{}^2\,\alpha_{02} + \alpha_{01}{}^2\,\alpha_{20} + 4\,\alpha_{10}\,\alpha_{01}\,\alpha_{11}) \\ &\quad + s_2\,(\alpha_{20}\,\alpha_{02} + 2\,\alpha_{10}\,\alpha_{12} + 2\,\alpha_{01}\,\alpha_{21} + 2\,\alpha_{11}{}^2) + s_1\,\alpha_{22}\,; \end{aligned}$$

and so on.

We define further

$$O^{hg}(u) = O^{10}\left(O^{h-1,\,g}(u)\right) = O^{01}\left(O^{h,\,g-1}(u)\right)$$

and establish without difficulty the formula

$$O^{hg}(uv) = \sum_{h_1=0}^{h} \sum_{g_1=0}^{g} \frac{h!}{h_1!\,(h-h_1)!} \frac{g!}{g_1!\,(g-g_1)!} O^{h_1 g_1}(u)\, O^{h-h_1, g-g}(v),$$

wherefrom we obtain

$$E\bar{\varphi}^{h+1}\bar{\psi}^{g} = O^{hg}(s_1\, \alpha_{10})$$

$$= \sum_{h_1=0}^{h} \sum_{g_1=0}^{g} \frac{h!}{h_1!\,(h-h_1)!} \frac{g!}{g_1!\,(g-g_1)!} O^{h_1 g_1}(s_1)\, O^{h-h_1, k-k_1}(\alpha_{10})$$

or

$$(34)\quad E\bar{\varphi}^{h+1}\bar{\psi}^{g} = \sum_{h_1=0}^{h} \sum_{g_1=0}^{g} \frac{h!}{h_1!\,(h-h_1)!} \frac{g!}{g_1!\,(g-g_1)!} \alpha_{h-h_1+1,\, g-g_1} E'\bar{\varphi}^{h_1}\bar{\psi}^{g_1}$$

where $E'\bar{\varphi}^{h_1}\bar{\psi}^{g_1}$ is to be received from $E\bar{\varphi}^{h_1}\bar{\psi}^{g_1}$ by replacing every factorial s_k in terms of $E\bar{\varphi}^{h_1}\bar{\psi}^{g_1}$ with s_{k+1}.

Finally, it is not difficult to write down the general expression for $E\bar{\varphi}^{h}\bar{\psi}^{g}$ in terms of α_{kl}.

Indeed, we write at first

$$\frac{\partial^{h+g}(u_1 u_2 \ldots u_s)}{\partial t^h\, \partial\theta^g} = \sum \frac{h!}{h_1!\ldots h_s!} \frac{g!}{g_1!\ldots g_s!} \frac{\partial^{h_1+g_1} u_1}{\partial t^{h_1}\, \partial\theta^{g_1}} \cdots \frac{\partial^{h_s+g_s} u_s}{\partial t^{h_s}\, \partial\theta^{g_s}},$$

where $u_1, u_2, \ldots, u_s$ are any functions of t, θ admitting the derivatives with regard to t and θ of all required orders and the summation is to be made for all positive integers $h_1, h_2, \ldots, h_s$ and $g_1, g_2, \ldots, g_s$ such that

$$h_1 + h_2 + \ldots + h_s = h, \quad g_1 + g_2 + \ldots + g_s = g.$$

Further we apply this formula to the relation (30) and, after some evident groupings of the terms of the received expression, we find lastly:

$$(35)\quad E\bar{\varphi}^{h}\bar{\psi}^{g} = \sum \frac{h!}{(\lambda_1!)^{\nu_1}\ldots(\lambda_r!)^{\nu_1}} \frac{g!}{(\mu_1!)^{\nu_1}\ldots(\mu_1!)^{\nu_r}} \frac{s_\nu}{\nu_1!\ldots\nu_r!} \alpha_{\lambda_1\mu_1}{}^{\nu_1}\ldots\alpha_{\lambda_r\mu_r}{}^{\nu_r}$$

where

$$(36)\qquad \nu = \nu_1 + \nu_2 + \ldots + \nu_r$$

and the sum must be taken for all positive integers λ_i, μ_i, ν_i, r such that

$$(37)\qquad \sum_{i=1}^{r} \nu_i\lambda_i = h, \quad \sum_{i=1}^{r} \nu_i\mu_i = g, \quad r \leq s,$$

whereas all pairs λ_i, μ_i must be different.

Take now q functions

$$\varphi_l(x) \quad (l = 1, 2, \ldots, q)$$

of our random variable x and their means

$$\bar{\varphi}_l = \frac{1}{s} \sum_{i=1}^{n} a_i \varphi_l(x_i)$$

for any sample S. It is evident how our previous considerations can be extended on this case, how in particular must be defined the operations $O^{10\cdots 0}, O^{010\cdots 0}, \ldots, O^{0\cdots 01}$ by means of which we could find different product moments of the means $\bar{\varphi}_l$. We shall omit this extension and only write down the general expression for the moments

$$E\bar{\varphi}_1^{h_1} \bar{\varphi}_2^{h_2} \ldots \bar{\varphi}_q^{h_q},$$

$h_1, h_2, \ldots, h_q$ being some positive integers. This expression is:

$$E\bar{\varphi}_1^{h_1} \bar{\varphi}_2^{h_2} \ldots \bar{\varphi}_q^{h_q} = \tag{38}$$

$$= \sum \frac{h_1!}{(\lambda_{11}!)^{\nu_1} \ldots (\lambda_{r1}!)^{\nu_r}} \cdots \frac{h_q!}{(\lambda_{1q}!)^{\nu_1} \ldots (\lambda_{rq}!)^{\nu_r}} \frac{s_\nu}{\nu_1! \ldots \nu_r!} \alpha_{\lambda_{11} \ldots \lambda_{1q}}^{\nu_1} \ldots \alpha_{\lambda_{21} \ldots \lambda_{rq}}^{\nu_r}$$

where

$$\nu = \nu_1 + \nu_2 + \ldots + \nu_r, \tag{39}$$

$$\alpha_{f_1 f_2 \ldots f_q} = s^{-(f_1 + f_2 + \ldots + f_q)} E\varphi_1^{f_1} \varphi_2^{f_2} \ldots \varphi_q^{f_q} \tag{40}$$

and the sum is extended on the all positive integer values of λ_{ij}, ν_i and r such that

$$\sum_{i=1}^{r} \nu_i \lambda_{ij} = h_j \ (j = 1, 2, \ldots, q), \ r \leq s, \tag{41}$$

whereas the indices

$$\lambda_{11} \lambda_{12} \ldots \lambda_{1q}, \ \lambda_{21} \lambda_{22} \ldots \lambda_{2q}, \ \ldots, \ \lambda_{r1} \lambda_{r2} \ldots \lambda_{rq}$$

of α's must be all different.

10. *Some applications.* — *a*) Putting

$$\varphi_1(x) = [\varphi(x) - \varphi_0]^a, \ \varphi_2(x) = [\varphi(x) - \varphi_0]^b,$$

where $\varphi_0 = E\varphi(x)$ and a, b are some real numbers, we shall have

$$\bar{\varphi}_1 = M[\varphi(x) - \varphi_0]^a, \ \bar{\varphi}_2 = M[\varphi(x) - \varphi_0]^b,$$

$$\alpha_{\lambda\mu} = s^{-\lambda-\mu} E(\varphi(x) - \varphi_0)^{\lambda a + \mu b}$$

and

$$E\bar{\varphi}_1^h \bar{\varphi}_2^g = E[M(\varphi - \varphi_0)^a]^h [M(\varphi - \varphi_0)^b]^g$$

$$= s^{-h-g} \sum \frac{h!}{(\lambda_1!)^{\nu_1} \ldots (\lambda_r!)^{\nu_r}} \frac{g!}{(\mu_1!)^{\nu_1} \ldots (\mu_r!)^{\nu_r}} \frac{s_\nu}{\nu_1! \ldots \nu_r!} [E(\varphi - \varphi_0)^{\lambda_1 a + \mu_1 b}]^{\nu_1} \ldots [E(\varphi - \varphi_0)^{\lambda_r a + \mu_r b}]^{\nu_r}$$

These moments can be applied to the research of the moments $E[M(x-\bar{x})^k]$ $(k=1,2,3,\ldots)$. Indeed, we have, putting $x_0 = Ex$:

$$M(x-\bar{x})^k = M[x-x_0-(\bar{x}-x_0)]^k$$
$$= \sum_{g=0}^{k} (-1)^g \frac{k!}{g!\,(k-g)!} M(x-x_0)^{k-g} (\bar{x}-x_0)^g$$
$$= \sum_{g=0}^{k} (-1)^g \frac{k!}{g!\,(k-g)!} M(x-x_0)^{k-g} [M(x-x_0)]^g$$

and accordingly

$$E[M(x-\bar{x})^k] = \sum_{g=0}^{k} (-1)^g \frac{k!}{g!\,(k-g)!} E[M(x-x_0)^{k-g}(M(x-x_0))^g]$$

We shall find the moments of the right side of this relation putting in the above written formulae

$$\varphi_1(x) = (x-x_0)^{k-g}, \quad \varphi_2(x) = x-x_0 = \varphi(x),$$
$$a = k-g, \quad b = 1, \quad h = 1, \quad g = g.$$

Then

$$\bar{\varphi}_1 = \mu'_{k-g(s)}, \quad \bar{\varphi}_2 = \bar{x}-x_0 = \mu'_{1(s)},$$
$$\alpha_{\lambda\mu} = s^{-\lambda-\mu} E(x-x_0)^{(k-g)\lambda+\mu} = s^{-\lambda-\mu} \mu_{(k-g)\lambda+\mu}$$

and, as it is easy to see,

(42) $$E[M(x-x_0)^{k-g}(M(x-x_0))^g] = E(\mu'_{k-g(s)}(\mu'_{1(s)})^g)$$
$$= s^{-g-1} \sum \frac{g!}{g_0!(g_1!)^{\nu_1}\ldots(g_r!)^{\nu_r}} \frac{s_\nu}{\nu_1!\ldots\nu_r!} \mu_{k-g+g_0} \mu_{g_1}^{\nu_1} \ldots \mu_{g_r}^{\nu_r},$$

where

$$\nu = 1+\nu_1+\nu_2+\ldots+\nu_r$$

and the sum is to be made for all different positive integers g_0, g_1, g_2, ... g_r and all positive (not necessary different) integers ν_1, ν_2, ... ν_r such that

$$g_0+\nu_1 g_1+\nu_2 g_2+\ldots+\nu_r g_r = g \quad (r \leq s)$$

The moments (42) can also be found by means of the operations O^{10} and O^{01}, the first of these operations being referred to zero differentiation with regard to t and the second = with regard to θ, whereas the generating function of the moments is

$$\left(\sum_{i=1}^{n} p_i e^{\frac{t}{s}(x_i-x_0)^{k-g}+\frac{\theta}{s}(x_i-x_0)}\right)^s$$

Denoting as above with O^{hg} the repetition of the operations O^{10} and O^{01} correspondingly h and g times we easily verify that

$$E\,[M\,(x-x_0)^{k-g}\,(M\,(x-x_0))^g] = O^{0g}\,(s_1\,\alpha_{10}),$$

where, accordingly to the general definition of $\alpha_{\lambda\mu}$ given above for our case,

$$\alpha_{10} = s^{-1}\,\mu_{k-g}.$$

Let us find, e. g., the moment $E\,[M\,(x-\bar{x})^3]$. We have:

$$\begin{aligned} E\,[M\,(x-\bar{x})^3] &= EM\,(x-x_0)^3 - 3\,E\,[M\,(x-x_0)^2\,M\,(x-x_0)] \\ &\quad + 3\,E\,[M\,(x-x_0)\,(M\,(x-x_0))^2] - E\,[M\,(x-x_0)]^3 \\ &= EM\,(x-x_0)^3 - 3\,E\,[M\,(x-x_0)^2\,M\,(x-x_0)] \\ &\quad + 2\,E\,[M\,(x-x_0)]^3; \end{aligned}$$

$$EM\,(x-x_0)^3 = O^{00}\,(s_1\,\alpha_{10}) = \mu_3 \qquad (k-g=3,\ g=0)$$

$$\begin{aligned} E\,[M\,(x-x_0)^2\,M\,(x-x_0)] &= O^{01}\,(s_1\,\alpha_{10}) \\ &= s_2\,\alpha_{10}\,\alpha_{01} + s_1\,\alpha_{11} = -s^2\,(s_2\,\mu_2\,\mu_1 + s_1\,\mu_3) \\ &= s^{-1}\,\mu_3 \qquad (k-g=2,\ g=1). \end{aligned}$$

$$\begin{aligned} E\,[M\,(x-x_0)]^3 &= O^{02}\,(s_1\,\alpha_{01}) = O^{01}\,(s_2\,\alpha_{01}^2 + s_1\,\alpha_{02}) \\ &= s_3\,\alpha_{01}^3 + 3\,s_2\,\alpha_{01}\,\alpha_{02} + s_1\,\alpha_{03} \\ &= s^{-3}\,(s_3\,\mu_1^3 + 3\,s_2\,\mu_1\,\mu_2 + s_1\,\mu_3) \\ &= s^{-2}\,\mu_3 \qquad (k-g=0,\ g=3) \end{aligned}$$

or

$$\begin{aligned} E\,[M\,(x-x_0)]^3 &= O^{02}\,(s_1\,\alpha_{10}) = s_3\,\alpha_{10}\,\alpha_{01}^2 + 2\,s_2\,\alpha_{01}\,\alpha_{11} \\ &\quad + s_2\,\alpha_{10}\,\alpha_{02} + s_1\,\alpha_{12} \\ &= s^{-3}\,(s_3\,\mu_1^3 + 3\,s_2\,\mu_1\,\mu_2 + s_1\,\mu_3) \\ &= s^{-2}\,\mu_3 \qquad (k-g=1,\ g=2) \end{aligned}$$

Finally

$$E\,[M\,(x-\bar{x})^3] = \mu_3 - \frac{3}{s}\,\mu_3 + \frac{2}{s^2}\,\mu_3 = \frac{(s-1)\,(s-2)}{s^2}\,\mu_3.$$

b) Let it be

$$\varphi_l\,(x) = x^{a_l} \quad (l = 1, 2, \ldots, q)$$

a_l being any real numbers. Then

$$\bar{\varphi}_l = Mx^{a_l} = m'_{a_l} \quad (l = 1, 2, \ldots, q)$$

$$\alpha_{f_1\,f_2\,\ldots\,f_q} = s^{-f}\,Ex^{f_1\,a_1 + \cdots + f_q\,a_q} = s^{-f}\,m_{f_1\,a_1 + \cdots + f_q\,a_q}$$

$$(f = f_1 + f_2 + \ldots + f_q)$$

and (38) gives the moments $E\,[(m'_{a_1})^{h_1}\ldots(m'_{a_q})^{h_q}]$ expressed in terms of m_h.

c) Putting

$$\varphi_l(x) = x^{a_l} - m_{a_l}, \quad m_{a_l} = Ex^{a_l} \quad (l = 1, 2, \ldots, q)$$

we shall have

$$\bar{\varphi}_l = m'_{a_l} - m_{a_l}, \; m'_{al} = Mx^{a_l} \quad (l = 1, 2, \ldots, q)$$

and

$$\alpha_{f_1 f_2 \ldots f_q} = s^{-\sum_{i=1}^{q} f_i} E[(x^{a_1} - m_{a_1})^{f_1} \ldots (x^{a_q} - m_{a_q})^{f_q}].$$

Then (38) will give us the moments

$$E[(m'_{a_1} - m_{a_1})^{h_1} \ldots (m'_{a_q} - m_{a_q})^{h_q}]$$

expressed in moments

$$E[(x^{a_1} - m_{a_1})^{f_1} \ldots (x^{a_q} - m_{a_q})^{f_q}]$$

d) Putting

$$\varphi_l(x) = (x - m_1)^{a_l}, \; m_1 = Ex \quad (l = 1, 2, \ldots, q)$$

we shall have

$$\bar{\varphi}_l = M(x - m_1)^{a_l} = \mu'_{a_l} \quad (l = 1, 2, \ldots, q)$$

$$\alpha_{f_1 f_2 \ldots f_q} = s^{-f} E[(x - m_1)^{f_1 a_1} \ldots (x - m_1)^{f_q a_q}]$$

$$= s^{-f} \mu_{f_1 a_1 + \ldots + f_q a_q}, \quad f = \sum_{i=1}^{q} f_i.$$

and (38) will give us the moments

$$E[(\mu'^{a_1})^{h_1} \ldots (\mu'_{a_q})^{h_q}]$$

expressed in moments μ_h.

e) Putting

$$\varphi_l(x) = (x - m_1)^{a_l} - \mu^a_l, \; \mu_{a_l} = E(x - m_1)^{a_l} \quad (l = 1, 2, \ldots, q),$$

we shall have

$$\bar{\varphi}_l = \mu'_{a_l} - \mu_{a_l}, \; \mu'_{a_l} = M(x - m_1)^{a_l} \quad (l = 1, 2, \ldots, q),$$

$$\alpha_{f_1 f_2 \ldots f_q} = s^{-\sum_{i=1}^{q} f_i} E[(x - m_1)^{a_1} - \mu_{a_1}]^{f_1} \ldots [(x - m_1)^{a_q} - \mu_{a_q}]^{f_q}$$

and (38) gives us the moments

$$E[(\mu'_{a_1} - \mu_{a_1})^{f_1} \ldots (\mu'_{a_q} - \mu_{a_q})^{f_q}].$$

11. *Functions of two and more random variables.* — We shall now consider many random variables and their functions beginning with the case of two variables x, y and one their function $\varphi(x, \nu)$.

Let the distribution $D(x, y)$ of x and y be represented by the table I where

$$x_1, x_2, \ldots, x_m \text{ and } y_1, y_2, \ldots, y_n$$

are the different values of x and y, p_{hk} is the

$y \backslash x$	x_1	x_2	$\ldots$	x_m	$p_{.k}$
y_1	p_{11}	p_{21}	$\ldots$	p_{m1}	$p_{.1}$
y_2	p_{12}	p_{22}	$\ldots$	p_{m2}	$p_{.2}$
$\vdots$	.	.	. . .	.	$\vdots$
y_n	p_{1n}	p_{2n}	$\ldots$	p_{mn}	$p_{.n}$
$p_{h.}$	$p_{1.}$	$p_{2.}$	$\ldots$	$p_{m.}$	1

TABLE I.

probability of the simultaneous equations $x = x_h$ and $y = y_k$ $(h = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n)$, $p_{h.}$ is the probability of $x = x_h$ $(h = 1, 2, \ldots, m)$ whatever value of y may be at the same time and p_k is the similar probability of $y = y_k$. It is evident that

$$\sum_{h=1}^{m} p_{hk} = p_{.k}, \quad \sum_{k=1}^{n} p_{hk} = p_{h.} \quad \sum_{h=1}^{m} \sum_{k=1}^{n} p_{hk} = \sum_{h=1}^{m} p_{h.} = \sum_{k=1}^{n} p_{.k} = 1.$$

Let us further consider some s observations which are independent one from another and during which the law $D(x, y)$ is inaltered. The results of these observations we can represent in the form of the table II, in

$y \backslash x$	x_1	x_2	$\ldots$	x_m	$a_{.k}$
y_1	a_{11}	a_{21}	$\ldots$	a_{m1}	$a_{.1}$
y_2	a_{12}	a_{22}	$\ldots$	a_{m2}	$a_{.2}$
$\vdots$	.	.	. . .	.	$\vdots$
y_n	a_{1n}	a_{2n}	$\ldots$	a_{mn}	$a_{.n}$
$a_{h.}$	a_1	$a_{2.}$	$\ldots$	$a_{m.}$	s

TABLE II.

which a_{hk} is the frequency of the pair $x = x_h$ and $y = y_k$ in s observations, $a_{h.}$ and $a_{.k}$ are correspondingly the frequencies of $x = x_h$ and $y = y_k$ in the same observations. This table is representing some sample of the size s. We shall denote this sample with S. We have evidently:

$$\sum_{h=1}^{m} a_{hk} = a_{.k}\,, \quad \sum_{k=1}^{n} a_{hk} = a_{h.}\,,$$

$$\sum_{h=1}^{m} \sum_{k=1}^{n} a_{hk} = \sum_{h=1}^{m} a_{h.} = \sum_{k=1}^{n} a_{.k} = s\,.$$

Beginning with one function of our two variables, $\varphi(x, y)$, we shall consider the moments of its mean

$$\bar{\varphi} = M\varphi(x, y) = \frac{1}{s} \sum_h \sum_k a_{hk}\, \varphi(x_h, y_k)$$

constructed for the values of x, y of the sample S.

It is not difficult to see that the function

$$u^s \equiv \left(\sum_{h=1}^{m} \sum_{k=1}^{n} p_{hk}\, e^{\frac{t}{s}\varphi(x_h, y_k)} \right)^s \tag{43}$$

is the generating function of the moments $E\,\bar{\varphi}^g$ $(g = 1, 2, 3, \ldots)$:

$$G(E\,\bar{\varphi}^g) = u^s. \tag{44}$$

Indeed, we have

$$u^s = \sum_{a_{hk}} \frac{s!}{a_{11}!\, a_{12}! \ldots a_{mn}!}\, p_{11}^{a_{11}} p_{12}^{a_{12}} \ldots p_{mn}^{a_{mn}}\, e^{\frac{t}{s}\sum_{h,k} a_{hk}\varphi(x_h, y_k)}$$

the sum $\sum_{a_{hk}}$ being taken for all positive integer a_{hk} such that $\sum_{h,k} a_{hh} = s$, $\sum_{h,k}$ denoting $\sum_{h=1}^{m} \sum_{k=1}^{n}$. This relation is equivalent to

$$u^s = \sum_{\bar{\varphi}} P_{\bar{\varphi}}\, e^{t\bar{\varphi}} \tag{45}$$

where

$$P_{\bar{\varphi}} = \frac{s!}{a_{11}!\, a_{12}! \ldots a_{mn}!}\, p_{11}^{a_{11}} p_{12}^{a_{12}} \ldots p_{mn}^{a_{mn}}$$

is evidently the probability of the sample S and, accordingly, of the corresponding mean $\bar{\varphi}$. Now (45) shows that (44) is true.

Comparing the present case with the case of a function of one variable we remark that all that has been said in §§ 1-6 can be repeated word in word and thus we shall obtain again

$$(46) \qquad E\overline{\varphi}^{g+1} = O\,(E\overline{\varphi}^{g})\,,$$

$$(47) \qquad E\overline{\varphi}^{g+1} = \sum_{f=1}^{g} \frac{g!}{f!\,(g-f)!}\,\alpha^{g-f+1}\,E'\overline{\varphi}^{f}\,,$$

$$(48) \qquad E\overline{\varphi}^{g} = \Sigma\,\frac{g!}{(\lambda_1!)^{\nu_1}\ldots(\lambda_r!)^{\nu_r}}\,\frac{s_\nu}{\nu_1!\ldots\nu_r!}\,\alpha_{\lambda_1}^{\nu_1}\ldots\alpha_{\lambda_r}^{\nu_r}\,,$$

where now

$$(49) \qquad \alpha_\lambda = s^{-\lambda}\,E\varphi^\lambda = s^{-\lambda}\sum_{h,\,k} p_{hk}\,[\varphi\,(x_h, y_k)]^\lambda\,.$$

Take further two functions $\varphi\,(x, y)$ and $\psi\,(x, y)$ of our random variables x, y and their means

$$\overline{\varphi} = \frac{1}{s}\sum_{h,\,k} a_{hk}\,\varphi\,(x_h\,,\,y_k)\,,\ \overline{\psi} = \frac{1}{s}\sum_{h,\,k} a_{hk}\,\psi\,(x_h\,,\,y_k)$$

for the sample S. We shall have now, as in § 9, the same operations O^{10} and O^{01} and the same formulae for the moments $E\overline{\varphi}^{g}\,\overline{\psi}^{l}$ as we have had there with the only difference that now

$$\alpha_{\lambda\mu} = s^{-\lambda-\mu}\sum_{h,\,k} p_{hk}\,[\varphi\,(x_h\,,\,y_k)]^\lambda\,[\psi\,(x_h\,,\,y_k)]^\mu$$

the general form of $\alpha_{\lambda\mu}$ being, however, the same as there, for we have now still

$$\alpha_{\lambda\mu} = s^{-\lambda-\mu}\,E\,\varphi^\lambda\,\psi^\mu$$

The relations (31), (31-*bis*), (34) and (35) still subsist.

It is clear that the considerations of § 9 are applicable to the case of many functions of x, y and that particularly we can write for them the formula (38) putting there

$$\alpha_{f_1 f_2 \ldots f_q} = s^{-\sum\limits_{i=1}^{q} f_i}\sum_{h,\,k} p_{hk}\,[\varphi_1\,(x_h\,,\,y^k)]^{f_1}\ldots[\varphi_q\,(x_h\,,\,y_q)]^{f_q}$$

In the same manner we can consider the case of many dependent random variables and their functions and we shall see that all formulae obtained for the case of functions of one variable will remain unaltered for this case the only difference consisting in the

new values of the quantities $\alpha_{f_1 f_2 \ldots f_q}$, their general definition being, however, the same in all cases:

$$\alpha^{f_1 f_2 \ldots f_q} = s^{-\sum_{r=1}^{q} f_i} E\, \varphi_1^{f_1} \varphi_2^{f_2} \ldots \varphi_q^{f_q},$$

whatever may be the variables for which are defined functions $\varphi_1, \varphi_2, \ldots, \varphi_q$.

Thus we see that the results obtained in the previous paragraphs are, *mutatis mutandis*, applicable to the case of any functions of any random variables. This shows the wide applications which may receive our formulae. Some of these applications we shall consider in the following paragraph.

We shall conclude this paragraph with a short indication of a theorem on the distribution of means $\overline{\varphi}$ and $\overline{\psi}$ of two functions $\varphi(x, y)$ and $\psi(x, y)$ of two random variables x, y whose distribution is defined by the table I. The demonstration of this theorem is quite similar to that of the theorem on the distribution of the means $\overline{x}$ and $\overline{y}$ of x and y themselves which the reader can find in my paper *Sur un théorème limite du calcul des probabilités* which will soonly be published in the « Recueil Mathématique de la Société Mathématique de Moscou ».

Theorem. Let $\overline{\varphi}$ *and* $\overline{\psi}$ *be the means of two functions* $\varphi(\mathrm{x}, \mathrm{y})$ *and* $\psi(\mathrm{x}, \mathrm{y})$ *of the dependent variables* x, y *whose distribution* $\mathrm{D}(\mathrm{x}, \mathrm{y})$ *is given by the table I and let it be*

$$\varphi_0 = E\varphi(x, y),\ \psi_0 = E\psi(x, y);\ \sigma_\varphi = \sqrt{E(\varphi - \varphi_0)^2},\ \sigma_\psi = \sqrt{E(\psi - \psi_0)^2}$$

$$r_{\varphi\psi} = \frac{E(\varphi - \varphi_0)(\psi - \psi_0)}{\sigma_\varphi \sigma_\psi}$$

Then, whatever real constants a *and* $\mathrm{b} > \mathrm{a}$, c *and* $\mathrm{d} > \mathrm{c}$ *may be, the probability of the simultaneous inequalities*

$$a < \frac{\overline{\varphi} - \varphi_0}{\sigma_\varphi \sqrt{2/s}} < b,\ c < \frac{\overline{\psi} - \psi_0}{6_\psi \sqrt{2/s}} < d$$

tends for $\mathrm{s} \to \infty$ *to the limit*

$$\frac{1}{\pi\sqrt{1 - r_{\varphi\psi_2}}} \int_a^b dt \int_c^d e^{-\frac{1}{1 - r_{\varphi\psi}^2}\left[t^2 - 2r_{\varphi\psi}\, t\theta - \theta^2\right]} d\theta .$$

It follows from this theorem that the distribution $D(\overline{\varphi}, \overline{\psi})$ of the means of our functions can be written approximately

$$D(\overline{\varphi}, \overline{\psi}) = \frac{s}{2\pi\sigma_\varphi\sigma_\psi\sqrt{1-r_{\varphi\psi}^2}} e^{-\frac{s}{2(1-r_{\varphi\psi}^2)}\left[\frac{(\overline{\varphi}-\varphi_0)^2}{\sigma_\varphi^2} - \frac{2r_{\varphi\psi}(\overline{\varphi}-\varphi_0)(\overline{\psi}-\psi_0)}{\sigma_\varphi\sigma_\psi} + \frac{(\overline{\psi}-\psi_0)^2}{\sigma_\psi^2}\right]},$$

if s is sufficiently great.

12. *Some applications.* — *a*) Put

$$\varphi(x, y) = (x - x_0)^a (y - y_0)_b, \quad x_0 = Ex, \quad y_0 = Ey,$$

a and b being some constant real numbers. Then

$$\overline{\varphi} = \frac{1}{s} \sum_{h,k} a_{hk} (x_h - x_0)^a (y_k - y_0)^b = \mu'_{ab(s)}$$

$$\alpha_\lambda = s^{-\lambda} E\varphi^\lambda = s^{-\lambda} \mu_{\lambda a, \lambda b},$$

if we denote $E(x - x_0)^{\lambda a} (y - y_0)^{\lambda b}$ with $\mu_{\lambda a, \lambda b}$, and we shall have

$$E(\mu'_{ab(s)})^g = s^{-g} \sum \frac{g!}{(\lambda_1!)^{\nu_1} \dots (\lambda_r!)^{\nu_r}} \frac{s_\nu}{\nu_1! \dots \nu_r!} \mu_{\lambda_1 a, \lambda_1 b}^{\nu_1} \dots \mu_{\lambda_r a, \lambda_r b}^{\nu_r}$$

Putting here $a = b = 1$ we shall obtain

$$E(\mu'_{11(s)})^g = s^{-g} \sum \frac{g!}{(\lambda_1!)^{\nu_1} \dots (\lambda_r!)^{\nu_r}} \frac{s_\nu}{\nu_1! \dots \nu_r!} \mu_{\lambda_1 \lambda_1}^{\nu_1} \dots {}_{\lambda_r \lambda_r}^{\nu_r}$$

b) Putting

$$\varphi(x, y) = x - x_0, \quad \psi(x, y) = y - y_0, \quad x_0 = Ex, \quad y_0 = Ey$$

we shall have

$$\overline{\varphi} = \overline{x} - x_0 = \mu'_{10(s)}, \quad \overline{\psi} = \overline{y} - y_0 = \mu'_{01(s)},$$

$$\alpha_{\lambda\mu} = s^{-\lambda-\mu} \mu_{\lambda\mu}$$

and

$$E(\mu'_{10(s)})^g (\mu'_{01(s)})^f =$$

$$= \sum \frac{g!}{(\lambda_1!)^{\nu_1} \dots (\lambda_r!)^{\nu_r}} \frac{f!}{(\mu_1!)^{\nu_1} \dots (\mu_r!)^{\nu_r}} \frac{s_\nu}{\nu_1! \dots \nu_r!} \mu_{\lambda_1 \mu_1}^{\nu_1} \dots \mu_{\lambda_r \mu_r}^{\nu_r}$$

c) Put

$$\varphi(x, y) = x - x_0, \quad \psi(x, y) = y - y_0, \quad \omega(x, y) = (x - x_0)(y - y_0),$$
$$x_0 = Ex, \quad y_0 = Ey.$$

Then

$$\overline{\varphi} = \mu'_{10(s)}, \quad \overline{\psi} = \mu'_{01(s)}, \quad \overline{\omega} = \mu'_{11(s)}$$

$$\alpha_{f_1 f_2 f_3} = s^{-f_1 f_2 f_3} E[(x - x_0)^{f_1} (y - y_0)^{f_2} (x - x_0)^{f_3} (y - y_0)^{f_3}]$$
$$= s^{-f_1 - f_2 - f_3} \mu_{f_1 + f_3, f_2 + f_3}$$

and

$$(50) \qquad E\left[(\mu'_{10\,(s)})^h \, (\mu'_{01\,(s)})^k \, (\mu_{11\,(s)})^l\right] =$$

$$= s^{-h-k-l} \sum \frac{h!}{(h_1!)^{\nu_1} \ldots (h_r!)^{\nu_r}} \, \frac{k!}{(k_1!)^{\nu_1} \ldots (k_r!)^{\nu_r}} \, \frac{l!}{(l_1!) \ldots (l_r!)^{\nu_r}} \cdot$$

$$\cdot \frac{s_\nu}{\nu_1! \ldots \nu_r!} \, \mu_{h_1+l_1,\, k_1+l_1}{}^{\nu_1} \ldots,\ \mu_{h_r+l_r,\, k_r+l_r}{}^{\nu_r},$$

where, as usually,

$$\nu = \nu_1 + \nu_2 + \ldots + \nu_r$$

and the sum is to be extended over all positive integer solutions of the equations

$$\sum_{i=l}^{r} \nu_i \, h_i = h, \quad \sum_{i=l}^{r} \nu_i \, k_i = k, \quad \sum_{i=l}^{r} \nu_i \, l_i = l \quad (r \leqq s)$$

giving different indices triplets

$$h_1,\ k_1,\ l_1;\ h_2,\ k_1,\ l_2;\ \ldots;\ h_r,\ k_r,\ l_r.$$

The moments just considered can be applied, for example, to the research of the standard deviation of the empirical product-moment

$$\overline{\mu}_{11} = \frac{1}{s} \sum_{h,\,k} a_{hk} \, (x_h - \overline{x}) \; y_k - \overline{y})$$

where the deviations of x_h and y_k are taken from their means $\overline{x}$ and $\overline{y}$. We have

$$\begin{aligned} \sigma_{\overline{\mu}_{11}}{}^2 &= E\,(\overline{\mu}_{11} - E\overline{\mu}_{11})^2 \\ &= E\overline{\mu}_{11}{}^2 - (E\overline{\mu}_{11})^2 \end{aligned}$$

and therefore we must find $E\overline{\mu}_{11}{}^2$ and $E\overline{\mu}_{11}$.

To find these moments we remark that

$$\begin{aligned} \overline{\mu}_{11} &= M\,(x - \overline{x})\,(y - \overline{y}) = M\left[(x - x_0) - (\overline{x} - x_0)\right]\left[(y - y_0) - (\overline{y} - y_0)\right] \\ &= M\,(x - x_0)\,(y - y_0) - (\overline{x} - x_0)\,(\overline{y} - y_0)\right] \\ &= \mu'_{11(s)} - \mu'_{10\,(s)}\,\mu'_{01\,(s)}. \end{aligned}$$

Then

$$(51) \qquad E\overline{\mu}_{11} = E\mu'_{11\,(s)} - E\mu'_{10\,(s)}\,\mu'_{01\,(s)},$$

$$E\,\overline{\mu}_{11}{}^2 = E\,(\mu'_{11\,(s)})^2 - 2\,E\,\mu'_{10\,(s)}\,\mu'_{01\,(s)}\,\mu'_{11\,(s)} + E\,(\mu'_{10\,(s)})^2\,(\mu'_{01\,(s)})^2$$

and we see that all moments on the right side of these relations are of the type (50).

But, in order to give an example of application of our algorithm to the present problem, we shall find the moments (51) by means of this algorithm.

We find:

$$
\begin{aligned}
E\,\mu'_{11\,(s)} &= s_1\,\alpha_{001} = \mu_{11}\,;\\
E\,\mu'_{10\,(s)}\,\mu'_{01\,(s)} &= O^{100}\,(E\,\mu'_{01\,(s)}) = O^{100}\,(s_1\,\alpha_{010})\\
&= s_2\,\alpha_{100}\,\alpha_{010} + s_1\,\alpha_{110}\\
&= s^{-2}\,(s_2\,\mu_{10}\,\mu_{01} + s\,\mu_{11}) = \frac{\mu_{11}}{s}\,;
\end{aligned}
$$

accordingly,

$$
E\,\overline{\mu}_{11} = \mu_{11} - \frac{\mu_{11}}{s} = \frac{s-1}{s}\,\mu_{11}\,.
$$

Further

$$
\begin{aligned}
E\,(\mu'_{11\,(s)})^2 &= O^{001}\,(s_1\,\alpha_{001}) = s_2\,\alpha_{001}{}^2 + s_1\,\alpha_{002}\\
&= s^{-2}\,(s_2\,\mu_{11}{}^2 + s_1\,\mu_{22})\\
&= \frac{s-1}{s}\,\mu_{11}{}^2 + \frac{1}{s}\,\mu_{22}\,;
\end{aligned}
$$

$$
\begin{aligned}
E\,\mu'_{10\,(s)}\,\mu'_{01\,(s)}\,\mu'_{11\,(s)} &= O^{110}\,(E\,\mu'_{11\,(s)}) = O^{100}\,O^{010}\,(s_1\,\alpha_{001})\\
&= O^{100}\,(s_2\,\alpha_{010}\,\alpha_{001} + s_1\,\alpha_{011})\\
&= s_3\,\alpha_{100}\,\alpha_{010}\,\alpha_{001} + s_2\,\alpha_{110}\,\alpha_{001} + s_2\,\alpha_{010}\,\alpha_{101}\\
&\quad + s_2\,\alpha_{100}\,\alpha_{011} + s_1\,\alpha_{111}\\
&= s^{-3}\,[s_2\,\mu_{10}\,\mu_{01}\,\mu_{11} + s_2\,\mu_{11}{}^2 +\\
&\quad + s_2\,\mu_{01}\,\mu_{21} + s_2\,\mu_{10}\,\mu_{12} + s_1\,\mu_{22}]\\
&= \frac{s-1}{s^2}\,\mu_{11}{}^2 + \frac{1}{s^2}\,\mu_{22}\,;
\end{aligned}
$$

$$
\begin{aligned}
E\,(\mu'_{10\,(s)})^2\,(\mu'_{01\,(s)})^2 &= O^{210}\,(E\,\mu'_{01\,(s)}) = O^{100}\,O^{100}\,O^{010}\,(s_1\,\alpha_{010})\\
&= O^{100}\,O^{100}\,(s_2\,\alpha_{010}{}^2 + s_1\,\alpha_{020})\\
&= O^{100}\,(s_3\,\alpha_{100}\,\alpha_{010}{}^2 + 2\,s_2\,\alpha_{010}\,\alpha_{110} + s_2\,\alpha_{100}\,\alpha_{020} + s_1\,\alpha_{120})\\
&= s_4\,\alpha_{100}{}^2\,\alpha_{010}{}^2 + s_3\,\alpha_{200}\,\alpha_{010}{}^2 + 2\,s_3\,\alpha_{100}\,\alpha_{010}\,\alpha_{110} +\\
&\quad + 2\,s_3\,\alpha_{100}\,\alpha_{010}\,\alpha_{110} + 2\,s_2\,\alpha_{110}{}^2 + 2\,s_2\,\alpha_{010}\,\alpha_{210} +\\
&\quad + s_3\,\alpha_{100}{}^2\,\alpha_{020} + s_2\,\alpha_{200}\,\alpha_{020} + s_2\,\alpha_{100}\,\alpha_{120}\\
&\quad + s_2\,\alpha_{100}\,\alpha_{120} + s_1\,\alpha_{220}\\
&= s^{-4}\,[s_4\,\mu_{10}{}^2\,\mu_{01}{}^2 + s_3\,\mu_{20}\,\mu_{01}{}^2 + s_3\,\mu_{02}\,\mu_{10}{}^2 +\\
&\quad + 4\,s_3\,\mu_{10}\,\mu_{01}\,\mu_{11} + 2\,s_2\,\mu_{11}{}^2 + 2\,s_2\,\mu_{01}\,\mu_{21} +\\
&\quad + 2\,s_2\,\mu_{10}\,\mu_{12} + s_2\,\mu_{20}\,\mu_{02} + s_1\,\mu_{22}]\\
&= \frac{2\,(s-1)}{s^3}\,\mu_{11}{}^2 + \frac{s-1}{s^3}\,\mu_{20}\,\mu_{02} + \frac{1}{s^3}\,\mu_{22}
\end{aligned}
$$

and therefore

$$E\bar{\mu}_{11}^2 = \frac{s-1}{s}\,\mu_{11}^2 + \frac{1}{s}\,\mu_{22} - \frac{2\,(s-1)}{s^3}\,\mu_{11}^2 - \frac{2}{s^2}\,\mu_{22} +$$

$$+ \frac{2\,(s-1)}{s^3}\,\mu_{11}^2 + \frac{s-1}{s^3}\,\mu_{20}\,\mu_{02} + \frac{1}{s^3}\,\mu_{22}$$

$$= \frac{s-1}{s}\left[1 - \frac{2}{s} + \frac{2}{s^2}\right]\mu_{11}^2 + \frac{(s-1)^2}{s^3}\,\mu_{22} + \frac{s-1}{s^3}\,\mu_{20}\,\mu_{02}$$

Thus $E\bar{\mu}_{11}$ and $E\bar{\mu}_{11}^2$ are found and we obtain finally

$$\sigma^2_{\bar{\mu}_{11}} = E\bar{\mu}_{11}^2 - (E\bar{\mu}_{11})^2$$

$$= \frac{(s-1)^2}{s^3}\,\mu_{22} - \frac{(s-1)\,(s-2)}{s^3}\,\mu_{11}^2 + \frac{s-1}{s^3}\,\mu_{20}\,\mu_{02}\,.$$

d) Put

$$\varphi\,(x, y) = (x - x_0)^2,\ \psi\,(x, y) = (y - y_0)^2,\ \omega\,(x, y) = (x - x_0)\,(y - y_0)$$

$$x_0 = Ex\,,\ y_0 = Ey\,.$$

Then

$$\bar{\varphi} = \bar{\mu}'_{20\,(s)}\,,\ \bar{\psi} = \mu'_{02\,(s)}\,,\ \omega = \mu'_{11\,(s)}$$

$$\alpha_{f_1 f_2 f_3} = s^{-f_1 - f_2 - f_3}\,\mu_{2f_1 + f_3,\ 2f_2 + f_3}$$

and

$$E\,(\mu'_{20\,(s)})^h\,(\mu'_{02\,(s)})^k\,(\mu'_{11\,(s)})^l =$$

$$= s^{-h-k-l}\sum \frac{h!}{(h_1!)^{\nu_1}\ldots(h_r!)^{\nu_r}}\ \frac{k!}{(k_1!)^{\nu_1}\ldots(k_r!)^{\nu_r}}\ \frac{l!}{(l_1!)^{\nu_1}\ldots(l_r!)^{\nu_r}}$$

$$\frac{s_\nu}{\nu_1!\ldots\nu_r!}\ \mu_{2h_1+l_1,\ 2k_1+l_1}^{\nu_1}\ \ldots\ \mu_{2h_r+l_r,\ 2k_r+l_r}^{\nu_r}$$

where the summation is to be made as in (50).

We shall not continue these examples. They are innumerable. It suffices to remark that allmost all problems of the theory of multiple correlation depending on the research of moments may be treated by our method.

13. *Functions of distributions changing from observation to observation.* — We shall conclude this paper by considering shortly the case when the distribution of the random variables whose functions we consider is not constant during all observations but is changing from observation to observation.

Let us begin with a function $\varphi\,(x)$ of one random variable x and suppose that its distribution $D_i\,(x)$ at the i^{th} observation is given by the table:

$$D_i\,(x):\ \begin{matrix} x_1^{(i)}, & x_2^{(i)}, & \ldots, & x_{n_i}^{(i)} \\ p^{(i)}, & p_2^{(i)}, & \ldots, & p_{n_i}^{(i)} \end{matrix}$$

where $x_h^{(i)}$ $(h = 1, 2, \ldots, n_i)$ are the different possible values of x at the i^{th} observation and $p_h^{(i)}$ $(h = 1, 2, \ldots, n_i)$ are their corresponding probabilities which satisfy the condition

$$\sum_{h=1}^{n_i} p_h^{(i)} = 1 .$$

We shall suppose, of course, that $\varphi(x)$ is defined and one valued for all values $x_h^{(i)}$ of x in the different distributions $D_i(x)$.

Let now be given any sample S of the size s

$$S: x', x'', \ldots, x^{(s)}$$

where x' is a value from the set $x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)}$; x'' is a value from the set $x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)}$; and so on. We shall construct the mean

$$\bar{\varphi} = \frac{1}{s}\left[\varphi(x') + \varphi(x'') + \ldots + \varphi(x^{(s)})\right]$$

and pose the problem of the research of the moments $E\bar{\varphi}^h$ $(h = 1, 2, 3, \ldots)$ of this mean.

It is very remarkable that this problem leads us to the same general formula for $E\bar{\varphi}^h$ as in § 4 on the condition to interprete it properly. This may be seen in the following manner.

It is clear at first that the product,

$$U(t) = \left(\sum_{h=1}^{n_1} p_h^{(1)} e^{\frac{t}{s}\varphi(x_h^{(1)})}\right)\left(\sum_{h=1}^{n_2} p_h^{(2)} e^{\frac{t}{s}\varphi(x_h^{(2)})}\right)\left(\sum_{h=1}^{n_s} p_h^{(s)} e^{\frac{t}{s}\varphi(x_h^{(s)})}\right)$$

is the generating function of the moments $E\bar{\varphi}^h$ so that

$$E\bar{\varphi}^h = \left[\frac{d^h U(t)}{dt^h}\right]_{t=0} .$$

If we consider further

$$E\bar{\varphi},\ E\bar{\varphi}^2,\ E\bar{\varphi}^3, \ldots$$

we shall easily remark that all these moments may again be written in the form

$$E\bar{\varphi} = s_1\,\alpha_1,\quad E\bar{\varphi}^2 = s_2\,\alpha_1^2 + s_1\,\alpha_2,\quad E\bar{\varphi}^3 = s_3\,\alpha_1^3 + 3\,s_2\,\alpha_1\,\alpha_2 + s_1\,\alpha_3$$

etc. on the condition that the quantities

$$s_1\,\alpha_1,\ s_1\,\alpha_2,\ s_1\,\alpha_3,\ s_2\,\alpha_1^2,\ s_2\,\alpha_1\,\alpha_2,\ s_3\,\alpha_1^3$$

etc., should be replaced with the sums

$$S_1\,\beta_1\,,\ S_1\,\beta_1^2\,,\ S_1\,\beta_1^3\,,\ S_2\,\beta_1\,\beta_2\,,\ S_2\,\beta_1\,\beta_2^2\,,\ S_3\,\beta_1\,\beta_2\,\beta_3$$

etc. where generally

$$\beta_\lambda^k = s^{-k}\,E_\lambda\,[\varphi(x)]^k$$

$$= s^{-k}\sum_{h=1}^{n\lambda} p_h^{(\lambda)}\,[\varphi(x_h^{(\lambda)})]^k$$

and the sum like $S_\nu\,\beta_1^{k_1}\,\beta_2^{k_2}\ldots\beta_\nu^{k_\nu}$ denotes the sum of all different products of ν factors like the product $\beta_1^{k_1}\,\beta_2^{k_2}\ldots\beta_\nu^{k_\nu}$ which can be made with the quantities $\beta_1, \beta_2, \ldots, \beta_s$, the indices $k_1, k_2, \ldots, k_\nu$ being the same in all products and conserving the same order in them.

We come in this way to the following general result which is not difficult to verify:

The moments $E\overline{\varphi}^h$ can be found one from another by the aid of the operation O defined in § 3 by means of the relation

$$E\overline{\varphi}^{h+1} = O\,(E\overline{\varphi}^h)$$

on the condition that in the final result the products like

$$s_\nu\,\alpha_{\lambda_1}^{\nu_1}\,\alpha_{\lambda_2}^{\nu_2}\ldots\alpha_{\lambda_r}^{\nu_r}\quad(\nu = \nu_1+\nu_2+\ldots+\nu_r\,,\ r\leqq s)$$

should be replaced with the corresponding sums

$$S_\nu\,\beta_1^{\lambda_1}\,\beta_2^{\lambda_1}\ldots\beta_{\nu_1}^{\lambda_1}\,\beta_{\nu_1+1}^{\lambda_2}\ldots\beta_{\nu_1+\nu_2}^{\lambda_2}\ldots\beta_{\nu-\nu_r+1}^{\lambda_r}\ldots\beta_\nu^{\lambda_r}$$

explained above.

This rule leads us immediately to another:

The moments $E\overline{\varphi}^h$ can be found from the general formula

$$E\overline{\varphi}^h = \sum \frac{h!}{(\lambda_1!)^{\nu_1}\ldots(\lambda_r!)^{\nu_r}}\,\frac{1}{\nu_1!\ldots\nu_r!}\,\cdot \tag{52}$$

$$\cdot\,S_\nu\,\beta_1^{\lambda_1}\ldots\beta_{\nu_1}^{\lambda_1}\,\beta_{\nu_1+1}^{\lambda_2}\ldots\beta_{\nu_1+\nu_2}^{\lambda_2}\ldots\beta_{\nu-\nu_r+1}^{\lambda_r}\ldots\beta_\nu^{\lambda_r}$$

where

$$\nu = \nu_1+\nu_2+\ldots+\nu_r$$

and the sum Σ must be taken for all positive integers λ_i, ν_i, r such that

$$\sum_{i=1}^{r}\nu_i\,\lambda_i = h\,,\ r\leqq s$$

and such that all λ_i are different.

Let us now remark that an operation like O can be defined which enables us to find the moments $E\overline{\varphi}^h$ one after another directly

without the intermediate expressions in α_i's. We shall denote this operation with H. It is defined by the conditions:

I. *The operation* H *must satisfy the addition and multiplication theorems of the differentiation of functions of one variable.*

II. *The operation* H *must satisfy the following special conditions:*

$$\begin{aligned} \beta_1^{k_1}\,\beta_2^{k_2}\ldots\beta_\nu^{k_\nu}\,H(S_\nu) &= S_{\nu+1}\,\beta_1\,\beta_2^{k_1}\,\beta_3^{k_2}\ldots\beta_{\nu+1}^{k_\nu}\,; \\ H(\beta_\lambda^{k}) &= \beta_\lambda^{k+1}\,; \\ H(A) &= 0\,, \end{aligned}$$

S_ν *being the sign of a sum described above,* ν_1, k_1, $k_2, \ldots, k_\nu$ *and* λ *being any positive integers and* A *a numerical constant.*

The operation H being thus defined we can find the moments $E\bar{\varphi}^k$ by the recurrent formula

$$E\bar{\varphi}^{k+1} = H(E\bar{\varphi}^k) \tag{53}$$

with the initial relation

$$E\bar{\varphi} = S_1\,\beta_1$$

We find, for example.

$$\begin{aligned} E\bar{\varphi}^2 &= H(E\bar{\varphi}) = H(S_1\,\beta_1) = \beta_1\,H(S_1) + S_1\,H(\beta_1) \\ &= S_2\,\beta_1\,\beta_2 + S_1\,\beta_1^2\,; \\ E\bar{\varphi}^3 &= H(E\bar{\varphi}^2) = H(S_2\,\beta_1\,\beta_2 + S_1\,\beta_1^2) \\ &= \beta_1\,\beta_2\;H(S_2) + S_2\,\beta_2\,H(\beta_1) + S_2\,\beta_1\,H(\beta_2) \\ &\quad + \beta_1^2\,H(S_1) + S_1\,H(\beta_1^2) \\ &= S_3\,\beta_1\,\beta_2\,\beta_3 + S_2\,\beta_2\,\beta_1^2 + S_2\,\beta_1\,\beta_2^2 + S_2\,\beta_1\,\beta_2^2 + S_1\,\beta_1^3 \\ &= S_3\,\beta_1\,\beta_2\,\beta_3 + 3\,S_2\,\beta_1\,\beta_2^2 + S_1\,\beta_1^3\,; \\ E\bar{\varphi}^4 &= H(S_3\,\beta_1\,\beta_2\,\beta_3 + 3\,S_2\,\beta_1\,\beta_2^2 + S_1\,\beta_1^3) \\ &= S_4\,\beta_1\,\beta_2\,\beta_3\,\beta_4 + 3\,S_3\,\beta_1\,\beta_2\,\beta_3^2 + 3\,S_3\,\beta_1\,\beta_2\,\beta_3^2 \\ &\quad + 3\,S_2\,\beta_1^2\,\beta_2^2 + 3\,S_2\,\beta_1\,\beta_2^3 + S_2\,\beta_1\,\beta_2^3 + S_1\,\beta_1^4 \\ &= S_4\beta_1\beta_2\beta_3\beta_4 + 6S_3\beta_1\beta_2\beta_3^2 + 3S_2\beta_1^2\beta_2^2 + 4S_2\beta_1\beta_2^3 + S_1\,\beta_1^4 \end{aligned}$$

and so on.

The application of the formula (53) is facilitated by the simple relation

$$\begin{aligned} &S_\nu\,\beta_{\nu_1+1}^{k_1}\,\beta_{\nu_1+2}^{k_2}\ldots\beta_\nu^{k_r-\nu_1}\;H(\beta_1^k\,\beta_2^k\ldots\beta_{\nu_1}^k) \\ &= \nu_1\,S_\nu\,\beta_1^k\,\beta_2^k\ldots\beta_{\nu_1-1}^k\,\beta_{\nu_1}^{k+1}\,\beta_{\nu_1+1}^{k_1}\ldots\beta_\nu^{k_\nu-\nu_1} \end{aligned} \tag{54}$$

which is almost evident.

We conclude this paragraph with two remarks.

1st *remark.* — It is clear that our formulae may be extended to the case when the distributions $D_i(x)$ are not more discrete as above but are defined in the most general manner as it was made in § 8.

2nd *remark.* — Let the distributions $D_i(x)$ $(i = 1, 2, \ldots, s)$ being defined in the most general manner are independent and such that the moments of some function $\varphi(x)$

$$\gamma_i = E_i \varphi(x), \; E_i |\varphi - \gamma_i|^{2+\delta}, \; E_i (\varphi - \gamma_i)^2 \; (i = 1, 2, \ldots, s)$$

are existing for some $\delta > 0$ and s however great and such that

$$0 < A \leqq E_i |\varphi - \gamma_i|^{2+\delta} \leqq B < \infty,$$
$$A \leqq E^i (\varphi - \gamma_i)^2 \leqq B$$
$$(i = 1, 2, 3, \ldots, s)$$

for s however great. Then, putting

$$u_i = \frac{\varphi(x^{(i)})}{s} \quad (i = 1, 2, \ldots, s),$$

where $x^{(i)}$ is a value of x observed in the i^{th} observation and belonging to the distribution $D_i(x)$, we shall have

$$\frac{1}{s^{\delta/2}} \frac{A}{B^{1+\delta/2}} \leqq \frac{\sum_{i=1}^{s} E|u_i - E u_i|^{2+\delta}}{\left[\sum_{i=1}^{s} E(u_i - E u_i)^2\right]^{1+\frac{\delta}{2}}} \leqq \frac{1}{s^{\delta/2}} \frac{B}{A^{1+\delta/2}},$$

wherefrom

$$\lim_{s \to \infty} \frac{\sum_{i=1}^{s} E|u_i - E u_i|^{2+\delta}}{\left[\sum_{i=1}^{s} E(u^i - E u_i)^2\right]^{1+\frac{\delta}{2}}} = 0.$$

Now $u_1, u_2, u_3, \ldots$ are independent,

$$\sum_{i=1}^{s} u_i = \frac{1}{s} \sum_{i=1}^{s} \varphi(x^{(i)}) = \overline{\varphi},$$

$$\sum_{i=1}^{s} E u_i = \frac{1}{s} \sum_{i=1}^{s} E_i \varphi(x)$$

and, thus, basing on the LIAPOUNOFF's theorem, we come to the following extension and generalisation of the theorem of § 6:

Theorem. If the distributions $D_i(x)$ *changing independently from observation to observation are such that all moments*

$$E_i\,\varphi(x),\ E_i\,|\,\varphi(x) - E_i\,\varphi(x)\,|^{2+\delta},\ E_i\,[\varphi(x) - E_i\,\varphi(x)]^2$$
$$(i = 1,\ 2,\ 3,\ \ldots)$$

of some function $\varphi(x)$ *are existing* δ *being some positive real number and are such that*

$$0 < A \leqq E_i\,|\,\varphi(x) - E_i\,\varphi(x)\,|^{2+\delta} \leqq B < \infty,$$
$$A \leqq E_i\,(\varphi(x) - E_i\,\varphi(x))^2 \leqq B$$
$$(i = 1,\ 2,\ 3,\ \ldots),$$

then the probability of the inequalities

$$z_1 < \frac{\bar{\varphi} - \frac{1}{S}\sum_{i=1}^{s} E_i\,\varphi(x)}{\sqrt{2\sum_{i=1}^{s} E_i\,[\varphi(x) - E_i\,\varphi(x)]^2}} < z_2$$

tends for $s \to \infty$ *to the limit*

$$\frac{1}{\sqrt{\pi}} \int_{z_1}^{z_2} e^{-z^2}\, dz$$

whatever real numbers z_1 *and* $z_2 > z_1$ *may be.*

14. *The case of two and more functions of one and more random variables.* — Remaining in the conditions of the preceding paragraph we shall now consider two functions $\varphi(x)$ and $\psi(x)$ of the variable x.

For this case we shall have the following results:

I. Let $\bar{\varphi} = M\varphi(x)$ and $\bar{\psi} = M\psi(x)$ be the means of $\varphi(x)$ and $\varphi(x)$ for the sample S considered in the preceding paragraph. Then the generating function for the moments $E\bar{\varphi}^h\,\bar{\psi}^g$ $(h,\ g = 1, 2, 3, \ldots)$ is

$$U(t, \theta) = \prod_{i=1}^{s} \left(\sum_{h=1}^{n_i} p_h^{(i)}\, e^{\frac{t}{s}\varphi(x_h^{(i)}) + \frac{\theta}{s}\psi(x_h^{(i)})} \right)$$

so that

$$E\varphi^h\,\bar{\psi}^g = \left[\frac{d^{h+g}\,U(t, \theta)}{dt^h\, d\theta^g}\right]_{t = \theta = 0}$$

II. We have

$$E\bar{\varphi}^h \bar{\psi}^g = H^{10} (E\bar{\varphi}^{h-1} \bar{\psi}^g) = H^{01} (E\bar{\varphi}^h \bar{\psi}^{g-1})$$

where H^{10} and H^{01} are the signs of the operations submitted to the general addition and multiplication rules of the differentiation with regard to two independent variables and to the following special rules:

$$\beta_1^{k_1 l_1} \ldots \beta_\nu^{k\nu\, l\nu} H^{10}(S_\nu) = S_{\nu+1}\, \beta_1^{10}\, \beta_2^{k_1 l_1} \ldots \beta_{+\nu\, 1}^{k\nu\, l\nu};$$
$$\beta_1^{k_1 l_1} \ldots \beta_\nu^{k\nu\, l\nu} H^{01}(S_\nu) = S_{\nu+1}\, \beta_1^{01}\, \beta_2^{k_1 l_1} \ldots \beta_{\nu+1}^{k\nu\, l\nu};$$
$$H^{10}(\beta_\lambda^{kl}) = \beta_\lambda^{k+1,l},\ H^{01}(\beta_\lambda^{kl}) = \beta_\lambda^{k,l+1};$$
$$H^{10}(A) = H^{01}(A) = 0\,.$$

In these relations k, l, ν, λ denote any positive integer numbers, β_λ^{kl} are defined by

$$\beta_\lambda^{kl} = s^{-k-l} E_\lambda [\varphi(x)]^k [\psi(x)]^l =$$
$$= s^{-k-l} \sum_{h=1}^{n\lambda} \rho_h^{(\lambda)} [\varphi(x_h^{(\lambda)})]^k [\psi(x_h^{(\lambda)})]^l\,,$$

S_ν has the same meaning as in the preceding paragraph and A is any numerical constant.

III. The general formula for $E\bar{\varphi}^h \bar{\psi}^g$ is

$$E\bar{\varphi}^h \bar{\psi}^g = \Sigma \frac{h!}{(\lambda_1 !)^{\nu_1} \ldots (\lambda_r !)^{\nu_r}} \frac{g!}{(\mu_1 !)^{\nu_1} \ldots (\mu_r !)^{\nu_r}} \frac{1}{\nu_1 ! \ldots \nu_r !}\,\cdot$$

$$\cdot S_\nu\, \beta_2^{\lambda_1 \mu_1} \beta_2^{\lambda_1 \mu_1} \ldots \beta_{\nu_1}^{\lambda_1 \mu_1} \beta_{\nu_1+1}^{\lambda_2 \mu_2} \ldots \beta_{\nu_1+\nu_2}^{\lambda_2 \mu_2} \ldots \beta_{\nu-\nu_r-1+1}^{\lambda_r \mu_r} \ldots \beta_\nu^{\lambda_r \mu_r}\,,$$

where

$$\nu = \sum_{i=1}^{r} \nu_i\,,\quad \sum_{i=1}^{r} \nu_i \lambda_i = h\,,\quad \sum_{i=1}^{r} \nu_i \mu_i = g\,,\quad r \leq s$$

and the sum Σ is to be made in the habitual manner.

It is evident how to proceed for the case of more than two functions of one random variable and we shall not develop this remark. We shall only state shortly the results concerning the case of one function of two stochastically dependent random variables.

We shall consider the distributions of two random variables x and y, the distributions which change independently one from

another as we go from an observation to another. Let such distribution for the i^{th} observation, $D_i\,(x, y)$, be defined by the table III, where all symbols

$y \backslash x$	$x_1^{(i)}$	$x_2^{(i)}$	$\dots$	$x_{m_i}^{(i)}$	$p_{.k}^{(i)}$
$y_1^{(i)}$	$p_{11}^{(i)}$	$p_{21}^{(i)}$	$\dots$	$p_{m_i 1}^{(i)}$	$p_{.1}^{(i)}$
$y_2^{(i)}$	$p_{12}^{(i)}$	$p_{22}^{(i)}$	$\dots$	$p_{m_i 2}^{(i)}$	$p_{.2}^{(i)}$
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\vdots$
$y_{n_i}^{(i)}$	$p_{1n_i}^{(i)}$	p_{2n_i}	$\dots$	$p_{m_i n_i}^{(i)}$	$p_{.n_i}^{(i)}$
$p_{h.}^{(i)}$	$p_{1.}^{(i)}$	$p_{2.}^{(i)}$	$\dots$	$p_{m_i.}^{(i)}$	1

TABLE III.

have the same meaning as in the table I of § 11 and the index i is placed to denote that all quantities entering in this table are defined for the i^{th} distribution $D_i\,(x, y)$.

Let now

$$x^{(i)}\,,\; y^{(i)}\,,\; (i = 1\,,\, 2\,,\, \dots,\, s)$$

be values of x, y observed in the i^{th} observation $(i = 1, 2, \dots, s)$. They constitute a sample of the size s which we shall denote again with S.

Consider further a function $\varphi\,(x\,,\, y)$ of these variables x, y and its mean

$$\bar{\varphi} = \frac{1}{s}\,[\varphi\,(x^{(i)},\; y^{(i)}) + \dots + \varphi\,(x^{(s)}y,\,{}^{(s)})\,]$$

based on the sample S. Then we shall have the following results:

I. The generating function of the moments $E\bar{\varphi}^g$ is given by

$$U\,(t) = \prod_{i=1}^{s} \left(\sum_{h=1}^{m_i} \sum_{k=1}^{n_i} p_{hk}^{(i)}\, e^{\frac{t}{s}\,\varphi\,(x_h^{(i)},\; y_k^{(i)})} \right)$$

so that

$$E\bar{\varphi}^g = \left[\frac{d^g\, U\,(t)}{dt^g}\right]_{t=0}.$$

II. The moments $E\bar{\varphi}^g$ ($g = 1, 2, 3, \ldots$) can be found one after another from the recurrent formula

$$E\bar{\varphi}^{g+1} = H\,(E\bar{\varphi}^g)$$

with the initial relation

$$E\bar{\varphi} = S_1\,\beta_1\,,$$

where H denotes the same operation as defined in the preceding paragraph and $\beta_\lambda{}^l$ are defined by

$$\begin{aligned}\beta_\lambda{}^l &= s^{-l}\,E_\lambda\,[\varphi\,(x)]^l \\ &= s^{-l} \sum_{h=1}^{m\lambda} \sum_{k=1}^{n\lambda} p_{hk}{}^{(\lambda)}\,[\varphi\,(x_h{}^{(\lambda)},\, y_k{}^{(\lambda)})]^l\,,\end{aligned}$$

λ and l being any positive integer numbers.

III. The general formula for the moments $E\bar{\varphi}^g$ is

$$E\bar{\varphi}^g = \Sigma\,\frac{g\,!}{(\lambda_1\,!)^{\nu_1}\ldots(\lambda_r)^{\nu_r}}\,\frac{1}{\nu_1\,!\ldots\nu_r\,!}\,.$$

$$S_\nu\,\beta_1{}^{\lambda_1}\,\beta_2{}^{\lambda_1}\ldots\beta_{\nu_1}{}^{\lambda_1}\,\beta_{1+\nu_1}{}^{\lambda_2}\ldots\beta_{\nu_1+\nu_2}{}^{\lambda_2}\ldots\beta_{\nu-{}_r+1}{}^{\lambda_r}\ldots\beta_\nu{}^{\lambda_r}\,,$$

where

$$\nu = \sum_{i=1}^{r} \nu_i\,,\quad \sum_{i=1}^{r} \nu_i\,\lambda_i = g\,,\quad r \leqq s$$

and the sum Σ must be taken in the habitual way.

These results are very important for the investigation of many problems concerning the theory of correlation for the distributions $D_i\,(x,\,y)$ considered here. The way, in which they can be extended for many functions of many random variables whose distributions vary independently from observation to observation, is obvious.

Tashkend, 5. II. 1928

A. DEGLI ESPINOSA

LA RICCHEZZA PRIVATA DEGLI ITALIANI NEL 1928

I.

È molto utile in genere conoscere la ricchezza privata di una nazione; ma è utile particolarmente per noi Italiani, in questa epoca che segna l'inizio di un periodo di lavoro, di lotta per arricchire, per conquistare quel posto nel mondo economico che il nostro giusto nazionalismo ci fa desiderare.

Liquidati, o quasi, gli inciampi creati dalla guerra e dalla politica di rivalutazione, il 1928 segna, appunto, con il suo finire, l'inizio di questo periodo di lavoro fecondo, ed è bene conoscere attualmente a quanto ammonti il nostro patrimonio, per poter nell'avvenire misurare il bilancio dei guadagni che allora saranno raggiunti.

Guardando il nostro passato, possiamo già dire che la ricchezza Italiana è cresciuta; questo è un buon augurio, ma non deve interessare quanto l'avvenire.

Sull'utilità specifica di questo calcolo si potrebbe osservare che esso rappresenta un duplicato, poichè, se non proprio per la stessa epoca, per una vicinissima, esattamente per il 1927, già altri calcoli sono stati eseguiti. Così quello del MORTARA. (1) Tale osservazione è però infondata, poichè, se è vero che un anno trascorso non varia la ricchezza di un paese, non bisogna dimenticare che il risultato di un calcolo sulla ricchezza, dipende dal metodo su cui esso è basato, ed è più o meno sicuro a seconda che il metodo adottato è più o meno rigoroso.

Nel nostro caso, per esempio, il risultato del calcolo del MORTARA, (1) così lontano da quello a cui noi giungiamo, ci fa pensare che

(1) MORTARA. *La ricchezza nazionale, il reddito nazionale e la pressione tributaria*, nel volume: « Movimento Economico Italiano – 1927 » della Banca Commerciale.

il metodo con cui è stato ottenuto, che l'autore non espone, debba essere differente dal nostro; non sarebbe possibile altrimenti spiegare quella forte differenza che presenta in confronto del nostro.

La convinzione nella giustezza dei criteri da noi seguiti, che sono quelli usati dal Prof. C. Gini nei suoi calcoli per il 1908, 1914, 1917, 1925, (1) giustifica quindi l'esistenza del presente studio.

In più il nostro risultato presenta, appunto per l'uguaglianza dei criteri seguiti nel calcolarlo, il vantaggio di essere perfettamente confrontabile con quelli del Prof. C. Gini e di formare, così, con essi una catena di cifre che misurano le variazioni e l'entità della ricchezza privata Italiana da un ventennio a questa parte. Dobbiamo però aggiungere che i nostri risultati sono stati quasi sempre ottenuti non solo aggiornando al 1928 i calcoli corrispondenti del Prof. C. Gini, ma anche mediante un calcolo diretto. La doppia valutazione costituisce la migliore garanzia della esattezza delle nostre conclusioni.

Non crediamo necessario aggiungere altre parole d'introduzione; le difficoltà e il significato di tali calcoli sono stati ampiamente illustrati da numerosi cultori di questa branca della statistica; i metodi da noi usati coincidono esattamente, ripetiamo, con quelli usati e illustrati dal Prof. Gini nella sua opera *L'ammontare e la composizione della ricchezza delle nazioni* e nei suoi scritti successivi sulla ricchezza delle nazioni.

A detta opera esauriente rimandiamo tutti coloro che desiderassero dimostrazioni e schiarimenti sui metodi usati e senz'altro passiamo all'esposizione del calcolo eseguito.

II.

Il metodo seguito nel presente calcolo, nel suo schema generale, è quello dell'inventario, criterio reale; quindi la ricchezza è stata suddivisa nelle varie categorie di beni che la compongono, e calcolata categoria per categoria.

I paragrafi seguenti contengono appunto l'esposizione di tali calcoli parziali, che in ultimo verranno riuniti e presentati in una tabella riassuntiva.

(1) C. Gini. *L'ammontare e la composizione della ricchezza delle nazioni*, Torino, Fr. Bocca, 1914. — *Problemi sociologici della guerra*, Bologna, Zanichelli, 1921. — *A comparison of the wealth and national income of several important nations* (*before and after the war*), Roma, Provv. Gen. dello Stato, 1925.

III.

VALORE DELLA PROPRIETÀ TERRIERA PRIVATA

1). — In base al valore venale medio per ettaro

In prima approssimazione, valuteremo la proprietà terriera nel 1928 basandoci sulle valutazioni antecedenti. Cioè, partendo dal valore venale medio per ettaro nel passato, abbiamo cercato di fissare quello attuale, tenendo conto del mutato potere d'acquisto della moneta.

Precisamente la base di partenza è la cifra data per il 1925 dal Prof. C. Gini. (1) Sarebbe stato preferibile staccarsi, per procedere nel calcolo, da una valutazione prebellica, poichè la stabilità dei prezzi e le statistiche più recenti davano alle valutazioni di prima della guerra una maggiore attendibilità di quella che offrano le valutazioni postbelliche, compiute in condizioni di forte variabilità dei prezzi e basate su statistiche, forse, ormai non più corrispondenti a realtà. (2)

Tuttavia abbiamo preferito attenerci alla valutazione del 1925, poichè le cambiate condizioni tributarie e di retribuzione del lavoro rendono impossibile stabilire una proporzione fra i prezzi del 1914 e quelli attuali.

È chiaro, infatti, come il passare da un prezzo di un'epoca trascorsa a quello attuale, tenendo conto della variazione del livello generale dei prezzi è legittimo e possibile soltanto se le altre circostanze che influiscono sulla dinamica dei prezzi del terreno si possono considerare invariate.

(1) C. Gini. *A comparison of the wealth and national income of several important nations, before and after the war.*

(2) C. Gini. *A comparison of the wealth*, ecc. pag. 6. L'A. espone dettagliatamente le circostanze per cui il risultato dei suoi calcoli è da considerarsi come largamente approssimativo. Precisamente: l'influenza delle rate d'interesse sui debiti, che faranno variare le tasse; l'esistenza di masse di carta moneta all'estero, che potranno essere assorbite dal Governo con debiti pubblici oppure dal mercato; la differenza dei prezzi dei beni venduti, che servono per la valutazione della ricchezza, e quelli dei beni non venduti, che costituiscono la maggior parte della ricchezza di una Nazione; la incertezza dei prezzi futuri; la mancanza di libertà nel fissare gli affitti delle case; la varietà della proporzione, fra regione e regione, con cui vengono compensati lavoro e capitale; le differenze tra i prezzi delle stesse merci da regione a regione.

L'impossibilità di passare per questa via dal valore dei terreni nel 1914 a quello dei terreni nel dopo guerra, si può controllare cercando di stabilire il prezzo venale medio per ettaro nel 1925, in base all'ipotesi che la variazione di tale prezzo sia proporzionale a quella del livello generale dei prezzi, e confrontando il prezzo così raggiunto con quello direttamente valutato dal Prof. C. GINI. (1)

Prima di passare alla descrizione delle operazioni compiute, riteniamo opportuno precisare il metodo con cui abbiamo fissato il prezzo medio per ettaro.

Il catasto agrario, ordinato con legge 21 luglio 1908, offre per il 1914 la suddivisione della superficie territoriale secondo la destinazione produttiva o improduttiva dei terreni: sarebbe quindi possibile per questo anno calcolare il prezzo medio per ettaro del terreno coltivato o non coltivato o sterile, escludendo dall'area totale quella delle acque e delle strade. Le notizie sulla distribuzione territoriale per gli anni successivi alla guerra non permettono invece di distinguere le aree occupate dalle acque e dalle strade, da quelle dei terreni sterili. Per ragioni di paragonabilità il prezzo medio per ettaro, di cui ci serviamo nei nostri calcoli, è, quindi, il rapporto fra il valore totale e la superficie produttiva, essendo tenuto conto del valore degli sterili (compreso nel valore totale) nel sovraprezzo, che risulta dall'averne trascurato la superficie, per ogni ettaro produttivo.

Accogliendo le valutazioni della ricchezza terriera Italiana presentate dal Prof. C. GINI, per gli anni 1914 (2) e 1925, (3) si è in grado di valutare con il criterio adottato il valore venale medio per ettaro negli anni stessi.

Prima di esporre le cifre è necessario compiere un'ultima osservazione. La superficie presa come denominatore della frazione, che rappresenta il prezzo cercato, comprende, oltre il terreno in mano dei privati, anche quello che non appartiene a questi; le valutazioni si riferiscono invece ai terreni di proprietà privata. È necessario quindi o innalzare il valore dei terreni per comprendervi anche quelli di proprietà pubblica, oppure detrarre dalla superficie quella porzione che si può ritenere di proprietà pubblica. L'una o l'altra via è indifferente: per maggior speditezza di calcolo abbiamo scelto la prima.

(1) C. GINI. *A comparison of the wealth* ecc., pag. 6.
(2) C. GINI. *Problemi sociologici della guerra*, pag. 256.
(3) C. GINI. *A comparison of the wealth* ecc., pag. 6.

Precisamente, basandoci sui dati per il 1908, ottenuti dal Prof. C. GINI (1), da cui risulta che il valore dei terreni dei privati sta a quello totale come 92 a 100, abbiamo corretto in proporzione le valutazioni degli anni da noi presi in considerazione. Una via diretta di calcolo era d'altra parte per questo particolare impedita dalla mancanza di materiale. (2)

I prezzi medi così calcolati sono i seguenti:

1914	L.	1764
1925	»	7875

Per verificare se, supponendo che il prezzo del 1928 stia a quello del 1925 come il livello generale dei prezzi di questo anno a quello del primo, si commette un'ipotesi troppo azzardata, sarà necessario studiare la dinamica della variazione del valore della terra dal 1914 al 1925. (3)

Le circostanze che possono aver influito sull'andamento del prezzo medio per ettaro sono: (4)

a) Il variare della distribuzione delle coltivazioni.

b) Il diminuito potere d'acquisto della moneta.

c) L'aumento del carico tributario.

d) La variazione della retribuzione del lavoro.

Cerchiamo ora di stabilire in che senso e in che misura possono, le circostanze elencate, aver influito sui prezzi.

La prima non può aver portato grandi conseguenze poichè l'aumento degli sterili dal 1914 al 1925 è lievissimo e poi è forse dovuto ad

(1) C. GINI. *L'ammontare e la composizione della ricchezza delle nazioni*, pag. 174. L'A. asserisce che la ricchezza terriera privata raggiunge i 37 miliardi mentre quella totale i 40 miliardi. Il rapporto è quindi come fra 92 e 100. Cioè l'8 % della ricchezza terriera è in mano delle persone giuridiche.

(2) Per il 1914 e 1917 sarebbe stato possibile, ma per il 1925 manca del tutto il materiale. Per il 1914 e 1917 si può vedere l'opera citata del GINI: *Problemi sociologici della guerra*, pagg. 257-263.

(3) Si potrebbe anche controllare il calcolo con i dati del 1917 forniti dal Prof. GINI in *Problemi sociologici della guerra*, pag. 247, ma il valore dei terreni in questo anno ha un carattere nettamente eccezionale. L'abbassamento di prezzo che si nota è dovuto al deperimento dei terreni a causa della meno intensa coltivazione, e alla distruzione dei vigneti e dei boschi, come dice il Prof. GINI stesso, ma queste particolari condizioni date dalla guerra, non hanno cambiato sostanzialmente le attitudini produttive del terreno.

(4) Per quanto è detto nella nota di sopra, non è stata presa in considerazione la circostanza: danni di guerra.

aumento di strade, fatto che, nel caso, potrebbe aver influito sul prezzo medio in senso contrario all'aumento degli sterili. Considerando la superficie agraria e forestale notiamo come la percentuale dei seminativi è andata diminuendo. Ciò può aver prodotto una diminuzione del prezzo medio per ettaro, ma essendo la variazione percentuale molto limitata, si può essere certi che per questa circostanza, ammesso che essa abbia potuto realmente influire sulla dinamica dei prezzi, la diminuzione del valore unitario medio del terreno deve essere stata trascurabile.

La tabella che segue dimostra quanto è stato ora detto:

Ripartizione della superficie territoriale (1)

Anni	Superficie territoriale	Seminativi	Superficie agraria e forestale: prati e pascoli	incolto produttivo	culture speciali (Vi., Ul., Fr.)	Boschi e castagneti	Totale superficie produttiva	Superficie improduttiva acque e strade
1914 (2)	28.661.037	13.233.000	6.080.000	1.035.000	1.486.000	4.564.000	26.398.000	2.263.000
	100	46.17	21.21	3.61	5.18	15.92	92.1	7.9
		50.13	23.23	3.92	5.63	17.29	100	
1925	31.013.713	13.227.000	6.853.400	1.251.000	1.525.600	5.585.400	28.442.700	2.571.013
	100	42.63	22.10	4.03	4.92	18.10	91.71	8.29
		46.50	24.09	4.40	5.36	19.63	100	
1927	31.013.713	13.009.979	6.973.800	1.251.000	1.626.521	5.585.400	28.446.700	2.567.013
	100	41.95	22.49	4.03	5.24	18.01	91.72	8.28
		45.73	24.52	4.40	5.72	19.63	100	

(1) I dati sono presi per il 1914 dall'« Annuario Statistico Italiano » del 1914, pag. 156. Per il 1917 dallo stesso annuario dell'anno 1919-21, pag. 206, e i dati del 1925-27 da quello del 1928, pag. 123.

(2) I dati di questa tabella hanno naturalmente subito delle lievi elaborazioni per poter confrontare la composizione territoriale nei vari anni, che nell'Annuario viene espressa per il 1925 e 1927 in modo differente dal 1914.

Il fattore *B*, la diminuzione del potere d'acquisto della moneta, preso isolatamente non ha bisogno di particolare commento: esso farebbe innalzare i prezzi dei terreni in proporzione della svalutazione della moneta stessa a parità di altre circostanze.

Ma in realtà è appunto l'uguaglianza delle altre circostanze che manca nella situazione in studio; siamo obbligati, quindi, ad analizzare l'influenza della diminuzione del potere d'acquisto assieme all'influenza di altri fattori, che ha inciso sull'opera della prima.

Tali fattori sono gli ultimi del nostro precedente elenco: la variazione della pressione tributaria e quella della retribuzione del lavoro.

È impressione generale che il carico tributario sia andato nel dopo guerra notevolmente crescendo, ma data la variazione del potere d'acquisto della moneta, che può mascherare quella della pressione fiscale, è bene analizzare a fondo l'argomento.

Vediamo i gettiti delle imposte e sovraimposte sui terreni:

GETTITO DELLE IMPOSTE E SOVRAIMPOSTE SUI TERRENI (1)

ANNI	IMPOSTA	SOVRAIMPOSTA	TOTALE
1914	84.637	196.828	281.465
1915	92.303	192.415	284.718
1916	99.864	209.526	309.390
1917	119.427	215.548	334.975
1918	129.871	217.049	346.920
1919	130.076	239.440	369.515
1920	129.767	290.483	420.250
1921	140.417	451.098	591.515
1922	151.920	657.637	809.557
1923	150.704	688.057	838.761
1924	154.227	753.873	908.100
1925	148.349	813.893	962.242
1926	148.671	896.970	1.045.641

(1) MINISTERO DELLE FINANZE. *La gestione delle imposte dirette dal 1914 al 1925*, pag. 89.
Per il 1926 « Annuario Statistico Italiano », 1928.

L'aumento del gettito, come osserva la Direzione Generale delle imposte dirette, in una sua pubblicazione del 1926 (1), è dovuto all'aumento delle aliquote poichè la materia imponibile, anche nella sua espressione catastale, è rimasta fino al 1924 inalterata. Tenendo conto delle sovraimposte locali la pressione tributaria appare nettamente cresciuta.

Citiamo a questo proposito i risultati di un calcolo del BERNARDINO (2) che mette a confronto il carico tributario del periodo 1911-13 con quello del periodo 1919-21.

Le imposte sui terreni pesavano nel periodo d'anteguerra per 3.06 per abitante, nel periodo posteriore alla guerra il carico medio per abitante raggiungeva 4.86. Le sovraimposte prima della guerra gravavano per 9.69 ogni abitante, dopo per 34.14.

È necessario poi tener presente che l'imposta e sovraimposta sui terreni non esaurisce il carico tributario su questi. Abbiamo infatti altre imposte che colpiscono direttamente il reddito dei terreni come l'imposta di ricchezza mobile sul reddito agrario (*Cat. B.*), l'imposta complementare progressiva sul reddito. (3)

Rimane ancora da notare che molte tasse senza essere direttamente gravanti sul prodotto della terra contribuiscono ad accrescere il carico tributario, fra queste per esempio è da porsi la tassa sugli scambi.

(1) MINISTERO DELLE FINANZE. *La gestione delle imposte dirette dal 1914 al 1925*, pag. 90.

(2) ANSELMO BERNARDINO. *La pressione tributaria dell'Italia e delle sue regioni prima e dopo la guerra*. Torino, Fratelli Bocca, 1928, pag. 21.

(3) Ecco l'elenco degli aumenti del carico tributario sul terreno:

Con Regio Decreto del 15 ottobre 1914, n. 1128, si elevò da 2 a 5 centesimi per lira, con effetto dal 1º gennaio 1915, l'addizionale a favore delle provincie e comuni terremotati.

Con legge 16 dicembre 1914, n. 1354, si aumentò di un decimo l'imposta principale sui terreni.

Nel 1917 questi aumenti furono riuniti in un ulteriore aumento dell'aliquota divenuta progressiva.

Infine con legge 16 ottobre 1924, alle cinque aliquote proporzionali venne sostituita un'unica aliquota del 10 % dell'estimo catastale che intanto veniva ad essere oggetto di una revisione.

Con Regio Decreto del 4 gennaio 1923, n. 16, il reddito ricavato dal proprietario del fondo, che lo coltiva in economia o colonia parziaria, venne assoggettato all'imposta di ricchezza mobile e compreso nella categoria *B*, a partire dal 1º gennaio 1923.

Con Regio Decreto del 30 dicembre 1923, n. 3062, fu istituita l'imposta complementare progressiva sul reddito, che a partire dal 1º gennaio 1925 colpisce il cumulo dei redditi.

Non sono sufficienti quindi, a dare un'idea della pressione fiscale, le cifre che misurano tale pressione esclusivamente riguardo alle imposte di carattere immobiliare.

Più espressivo è il rapporto fra pressione tributaria generale e ricchezza.

L'onere tributario ragguagliato alla ricchezza privata era 4.08 prima della guerra, nel 1920 era 10.77. (1) Rimane quindi chiaro che il carico tributario è fortemente aumentato nel dopoguerra.

Rimane ora da vedere in che direzione questo aumento può avere influito sul prezzo dei terreni stessi.

Parlando di prezzi occorre distinguere fra prezzi di consumo e prezzi dei beni strumentali e capitali. Il prodotto della terra e quindi il rendimento netto, quando dal primo si detraggono i carichi di qualsiasi specie, viene misurato dal prezzo dei beni di consumo, mentre il valore del fondo viene fissato capitalizzando il rendimento netto.

Rimanendo costante il livello dei prezzi dei beni di consumo, è chiaro che il prezzo dei beni capitali diminuirà tanto più quanto più forti saranno i carichi che gravano il prodotto. In regime di prezzi crescenti avremo, quando si supponga che la causa che fa aumentare i prezzi sia indipendente dall'aumento del carico che grava il prodotto, che il capitale crescerà di prezzo tanto meno, in confronto dell'aumento dei prezzi dei beni di consumo, quanto più crescerà il carico da detrarre dal prodotto per avere il rendimento netto. (2)

Nel nostro caso possiamo quindi concludere che i prezzi dei terreni, per effetto della svalutazione della moneta sono cresciuti, ma per effetto dell'aumento del carico tributario sono cresciuti meno, in confronto dell'anteguerra, dei prezzi dei beni di consumo.

Dobbiamo infine analizzare l'azione della svalutazione della moneta unitamente a quella dell'aumento della retribuzione del lavoro. Dobbiamo cioè indagare se i salari reali sono cresciuti in modo tale da alterare la distribuzione prebellica del reddito.

(1) A. Bernardino. *Op. cit.* rispettivamente a pagine 23, 57, 98, 99.

(2) Vedi al proposito C. Gini *Problemi sociologici della guerra*, pag. 300. L'A. distingue i prezzi dei beni capitali e strumentali dai « prezzi delle derrate e degli altri beni di uso diretto che costituiscono il fine della produzione ». Egli dice « le inevitabili asprezze delle imposte del dopo guerra in confronto dell'anteguerra, diminuendo il rendimento netto dei beni capitali e strumentali, tenderà a far sì che il prezzo loro, in confronto dell'anteguerra, risulti meno elevato di quello delle derrate e dei beni di consumo diretto ».

Si capisce che una tale alterazione influisce sul valore dei terreni in quanto che, secondo che essa è verificata a favore dei proprietari o dei salariati, detto valore verrà ad essere innalzato o diminuito, relativamente al livello che avrebbe raggiunto se la distribuzione fosse rimasta inalterata. In sostanza si tratta di fissare se l'aumento dei salari è avvenuto in misura proporzionale o più che proporzionale all'aumento del livello generale dei prezzi.

Considerando vari tipi di economia, vediamo che nella Provincia di Treviso, nella conduzione a mezzadria, prima della guerra la ripartizione del prodotto ne dava al proprietario il 35 %; dopo la guerra, nel 1920-21, la percentuale era scesa al 14,5 %. Nella conduzione ad affitto, figura in cui il proprietario assume l'aspetto d'imprenditore, la riduzione è minore ma sempre sensibile, precisamente dal 40 % al 30 %. (1) Questa alterazione nella distribuzione si è attualmente smorzata, ma tuttavia ancora, sebbene leggermente, fa sentire la sua influenza.

Alcuni altri dati sui vari tipi di aziende agrarie dimostreranno esaurientemente lo spostamento delle proporzioni fra retribuzione del lavoro e reddito del proprietario che è avvenuto dal periodo prima della guerra a quello seguente. La seguente tabella dà i numeri indici che descrivono le variazioni del reddito del contadino e del proprietario che si sono avute nel Friuli: (2)

Numeri indici del reddito della proprietà terriera nel Friuli

Anni	Coltivatore	Proprietario
1914	100	100
1915	96	98
1916	84	80
1917	80	75
1918	167	133
1919	188	162
1920	235	195
1921	356	340
1922	444	353
1923	457	451
1924	478	395

Ancora altri indici mostrano il movimento in studio:

(1) C. Gini. *Corso di Sociologia*, Anno Acc. 1926-27, Università di Roma. Castellani, pag. 347-348.

(2) I dati sono stati forniti dal Prof. Gaetano Pietra, ordinario di Statistica all'Università di Padova.

INDICI DI AUMENTO DELLA RIMUNERAZIONE DEL LAVORO MANUALE E DI AMMINISTRAZIONE E DEL REDDITO FONDIARIO DALL'ANTEGUERRA AL DOPOGUERRA PER ALCUNI TIPI DI AZIENDE ITALIANE (1)

TIPI DI AZIENDE STUDIATI	Indici di aumento		
	della retribuzione del lavoro		del reddito fondiario
	manuale	di amministrazione	
Sistemi intensivi			
Tipi di aziende irrigue lombardo-piemontesi:			
Aziende irrigue a prevalente produzione di latte e riso	8,4	4,4	2,6
Aziende irrigue a prevalente produzione di latte.	..	..	..
Aziende irrigue a prevalente produzione di riso .	..	..	..
Tipi di aziende asciutte emiliane con colture industriali:			
Azienda bolognese con viticoltura	4,4	3,8	4
Azienda romagnola con frutticoltura	5,8	4,1	6,5
Tipi di aziende monferrine con coltura speciale di vite:			
Piccola azienda a mezzadria	6,5	3,8	7,3
Azienda ad economia diretta	5,7	4,8	5,6
Tipi di aziende a coltura promiscua dell'Italia centrale:			
Azienda toscana di pianura con coltura di vite	5,8	5	5,2
Azienda litoranea toscana con coltura di olivo.	5,3	4,5	4,8
Azienda umbra di piano con coltura di vite	5,5	4,8	5
Azienda umbra collinare con coltura di olivo.	5,5	4,8	4,8
Tipi di aziende a colture legnose speciali:			
Azienda viti-olivicola pugliese	4,7	4,1	6,6
Piccola azienda viticola siciliana	5,9	3,5	3,7
Azienda agrumicola: limoneto.	4,7	2,1	0,9
Azienda agrumicola: aranceto.	..	..	..
Sistemi estensivi			
Tipi di aziende latifondistiche a prevalente cerealicoltura:			
Azienda del Tavoliere di Puglia	5	4,8	2,4
Altro tipo di azienda del Tavoliere	5,4	5,1	2
Tipi di aziende latifondistiche cerealicole-pastorali:			
Azienda del latifondo siciliano	3,7	4	3,4

(1) La presente tabella è stata costruita con i dati forniti dall'opera di G. TASSINARI, *Saggio intorno alla distribuzione del reddito nell'Agricoltura Italiana*. Il periodo postbellico si riferisce agli anni 1921-22.

La maggioranza degli indici di variazione calcolati dal TASSINARI, facendo uguale ad 1 il valore prebellico, per la rimunerazione del lavoro manuale o d'amministrazione e del reddito fondiario mostrano come la prima sia aumentata più dell'ultimo. (1)

Per il nostro scopo occorre considerare questi indici di aumento in confronto della variazione dei prezzi.

Nel 1922 il livello generale dei prezzi era 507.5 essendo uguagliato a 100 il livello dei prezzi nel 1913, il che equivale a un indice di variazione di 5.07 avendo fatto il livello dei prezzi nel 1913 uguale a 1, per poter confrontare la variazione dei prezzi con quella della rimunerazione del lavoro e della proprietà.

Vediamo allora che su 15 tipi di aziende presi in considerazione dal TASSINARI, e di cui abbiamo mostrato gli indici di variazione relativi ai fattori in questione, 4 soltanto mostrano che il reddito fondiario è aumentato più di quello che siano aumentati i prezzi. Considerando la retribuzione del lavoro manuale notiamo che questa per 11 su 15 è cresciuta più dei prezzi, in modo nettissimo. Riguardo alla retribuzione del lavoro amministrativo e di direzione vediamo che essa è aumentata più di quello che siano aumentati i prezzi soltanto in due casi su 15.

Interpretando queste conclusioni, possiamo dire che l'aumento dei prezzi è stato compensato per i lavoratori dall'aumento dei salari nominali, anzi più che compensato.

Per il proprietario invece, e per l'amministratore, che spesso costituisce un secondo aspetto del primo, l'aumento dei prezzi rappresenta una perdita, cioè una diminuzione delle entrate reali.

Adattando queste conclusioni alla nostra indagine, possiamo dire che la variazione della distribuzione è stata sfavorevole alla proprietà terriera in quanto che la retribuzione del lavoro è aumentato più che proporzionalmente all'aumento del livello generale dei prezzi, mentre meno che proporzionalmente è avvenuto l'aumento del reddito fondiario.

Da ciò si deduce, che il prezzo venale medio per ettaro, deve essere cresciuto meno che proporzionalmente all'aumento del livello generale dei prezzi.

(1) È questo un fenomeno generale: nel « The Economic Journal » del dicembre 1928 Mr. J. A. VENN, nel suo articolo *The incidence of taxation in agriculture*, dà alcuni dati che illustrano il fenomeno in questione per l'Inghilterra. Nel periodo 1884-93, il prodotto del fondo si ripartiva in questo modo fra lavoratori e proprietario: 17,3 % andava al secondo, il 20,5 % ai primi; nel 1919 rispettivamente il 10,0 per cento e il 24,0 %, nel 1926 l'11,6% e il 28,9%.

Le considerazioni fatte spiegano esattamente come il valore della ricchezza terriera in Italia, calcolato in base all'ipotesi che il prezzo venale medio per ettaro fosse aumentato nella stessa proporzione del livello generale dei prezzi, avrebbe dato per il 1925 una cifra molto superiore al vero. (1) Per queste stesse ragioni non ci fu possibile, per calcolare la ricchezza terriera privata nel 1928 in base all'ipotesi suesposta, partire dalla valutazione del 1914, ma dovemmo invece basarci su quella del 1925, che forse presenta una minore esattezza della prima.

Nel passare dal 1925 al 1928, i fattori analizzati or ora hanno avuto alcuni un' influenza contraria a quella avuta fino al 1925, altri non hanno più avuto influenza.

La diminuzione dei prezzi iniziatasi nel 1927, ha indubbiamente portato anche all'agricoltura delle condizioni di vita difficili e faticose, che hanno forse minacciato di far scendere i prezzi della terra più di quello che scendessero quelli dei beni di consumo, ma d'altra parte relativamente alla situazione del 1925, anche la retribuzione del lavoro è diminuita più di quello che sia diminuito il livello generale dei prezzi.

Questi due fattori, agenti in senso contrario, fanno ritenere che all'ingrosso le variazioni del prezzo medio venale per ettaro siano state, dal 1925 al 1928, proporzionali a quelle del livello generale dei prezzi.

Passiamo ora, dopo questa lunga giustificazione, al calcolo.

Nel 1925 il Prof. C. Gini valutava a 200 miliardi la ricchezza terriera privata compresa negli antichi confini. Supponendo che la ricchezza delle nuove provincie avesse la stessa composizione percentuale di quella dell'Italia, possiamo portare la cifra del Gini a 208 miliardi per comprendervi anche i terreni delle terre redente. (2) Infine supponendo che la ricchezza terriera nazionale stesse a quella privata nella stessa proporzione in cui stava nel 1908, (3) abbiamo 224 miliardi come ricchezza nazionale terriera nel 1925. Dividendo questa cifra per la superficie agraria e forestale, valutata in 28.442.700 ettari (4) concludiamo che il prezzo venale medio per ettaro era nel 1925 di lire 7875.

(1) Ponendo il livello dei prezzi nel 1914 = 100, quello del 1925 risulta = 660,6 (serie Bachi), il prezzo medio per ettaro che era nel 1914 di lire 1764, nel 1925 avrebbe dovuto essere di lire 11.653, mentre calcolandolo in base alla valutazione del Prof. Gini, con gli stessi criteri adoprati per il 1914, risulta di lire 7875.

(2) C. Gini. *A comparison of the wealth* ecc., pag. 6.

(3) Vedi pag. 295.

(4) « Annuario Statistico Italiano », anno 1928.

Nello stesso anno il livello generale dei prezzi era 646,24; (1) nel 1928, era, invece, 491.36.

Con una semplice proporzione troviamo che il prezzo medio per ettaro era nel 1928 di lire 5985. Questa cifra moltiplicata per la superficie totale agraria e forestale corrispondente al 1927, (2) che noi possiamo estendere al 1928, dà la valutazione delle terre italiane per il 1928, in lire 170 miliardi. Tale valore ridotto nella stessa proporzione del 1908 ci dà il valore della ricchezza terriera privata in 151 miliardi di lire.

2). — In base al reddito netto

Sarà utile compiere una riprova del calcolo precedente mediante la capitalizzazione del reddito netto. Per questo si deve naturalmente determinare il valore della produzione lorda. Vari autori hanno compiuto tale valutazione nel dopo guerra. Ricorderemo fra le prime quella del Prof. Michele Carlucci. (3)

Egli giungeva al risultato che nel trennio 1920-22, la produzione annua media, ai prezzi di allora, della nostra agricoltura andava sui 36 (35.600 milioni) miliardi sul campo, e sui 40 miliardi sul mercato di vendita.

Il Bordiga, calcolava nella stessa epoca, (4) in base alla revisione degli estimi avvenuta nel 1924, che la produzione agricola superava i 36 miliardi

Una terza valutazione è quella dell'Ing. Zattini. Egli si riferisce alla produzione media nel periodo che decorre dal 1909, cioè dall'epoca in cui sorse la statistica agraria, (5) e che si arresta nel 1925. Dobbiamo osservare a questo proposito, che tale media, data la depressione causata nel periodo della guerra dalla minore cura della coltivazione, nel periodo immediatamente successivo alla guerra dalle agitazioni sovversive, nonostante il risollevamento a partire dal 1923, viene a raffigurare una valutazione inferiore a quella vera per il 1928.

(1) Serie calcolata dal Consiglio Provinciale dell'Economia di Milano.

(2) Vedi pag. 297.

(3) L. Maroi. *Valore e reddito della terra*, Macerata, 1929. A pag. 106 e seguito, viene citato il calcolo del Carlucci *Valore attuale dell'agricoltura italiana*, 30 giugno 1923.

(4) Id. id. pag. 114, viene citato il calcolo del Bordiga.

(5) Ing. G. Zattini. *Valutazione della produzione lorda dell'agricoltura italiana*, Roma, 1925.

D'altra parte l'aumento della produzione dal 1923 in poi non è a carattere transitorio. Ne consegue che la media sopradetta è, ai nostri scopi, priva di significato. Per avere una valutazione della produzione agricola nel giro di tempo che va dal 1924 al 1928, bisogna riferirsi alla produzione di questi ultimi anni.

Per tali ragioni si deve considerare la valutazione dello ZATTINI inferiore alla realtà.

La cifra a cui egli giunge è ai prezzi dell'anteguerra di 8 miliardi e 245.700.000. (1)

Il PORRI ha compiuto nel 1926 una nuova valutazione, che ha confermata l'opinione che egli aveva riguardo alla valutazione dello ZATTINI, che riteneva troppo bassa, ed è giunto alla cifra di 50 miliardi. (2)

Infine il FRANCIOSA ha valutato la produzione della nostra agricoltura nel 1927, (3) giungendo alla cifra di 42.408 milioni. Una prima ragione della differenza dei risultati di questi due ultimi autori si trova nel fatto che il PORRI ha calcolato il valore realizzato sui mercati di vendita, (4) mentre il FRANCIOSA ha valutato quello sul campo.

Bisogna notare poi, come osserva il FRANCIOSA, che l'anno 1927 è stato particolarmente sfortunato per l'agricoltura, sia a causa di cattive condizioni climatiche sia per la caduta dei prezzi. (5)

Il MAROI, criticando le valutazioni ora esposte, dice che quella del PORRI pecca per eccesso, (6) ammette però che nella nostra agricoltura ci sia stato un progresso. Egli per il 1927 dichiara attendibile la valutazione del FRANCIOSA in 42 miliardi e mezzo.

Naturalmente tale cifra si concilia perfettamente, nonostante la caduta dei prezzi, con quella dello ZATTINI, poichè questa, come si è detto, deve essere errata per difetto.

La cifra del FRANCIOSA mostra come la nostra produzione sia aumentata dal 1920-21. L'aumento, d'altra parte, nota il MORTARA, (7)

(1) G. ZATTINI. *Op. cit.*, pag. 172.

(2) L. MAROI. *Op. cit.*, pag. 132, viene citato il calcolo del PORRI.

(3) L. FRANCIOSA. *La produzione agraria ed il suo valore nell'anno* 1927. « Economia », n. 9, anno VI.

(4) L. MAROI. *Op. cit.*, pag. 141.

(5) L. FRANCIOSA. Articolo citato.

(6) L. MAROI. *Op. cit.*, pag. 156.

(7) G. MORTARA. *La ricchezza nazionale, il reddito*, ecc., pag. 389.

risalta da molti indici come l'estensione del terreno produttivo, l'aumento degli edifici rustici, del corredo di strumenti agricoli e macchine, del capitale rappresentato da bestiame, dei canali d'irrigazione. Il MAROI osserva poi come un indice di miglioramento sia anche la trasformazione del tipo di proprietario da aristocratico a piccolo borghese attivo e produttivo. (1)

Noi dobbiamo vedere, ora, se la valutazione del FRANCIOSA è adottabile per il nostro calcolo.

Diciamo subito che essa si adatta benissimo, dal punto di vista del metodo con cui è stata calcolata.

Infatti a noi occorre il valore della produzione sul campo, che è appunto il valore realizzato dall'agricoltura. Quell'aumento che si nota dal valore sul campo a quello sul mercato di vendita rappresenta un'entrata corrispondente ad altre attività (trasporti, senserie, ecc.) e quindi da capitalizzarsi in altra parte di questo calcolo. Così facendo si viene ad eliminare la ripetizione inevitabile a cui si andrebbe incontro valutando la produzione sul mercato.

Rimane sempre però da risolvere la questione se si può estendere il valore della produzione dal 1927 al 1928.

Si capisce che il valore della proprietà terriera non varia necessariamente da anno ad anno a seconda che la produzione sia abbondante o scarsa. Tale valore può però variare secondo la natura della variazione del valore della produzione. Occorre distinguere infatti riguardo a queste variazioni alcuni differenti aspetti che esse possono assumere. È possibile, per comodità, accettare la terminologia economica del movimento secolare ed accidentale. Si dirà che una variazione della produzione è corrispondente ad un movimento secolare, quando essa ha un carattere di trasformazione permanente; in altre parole quando si tratta di un miglioramento o di un peggioramento che diverrà lo stato ordinario dell'agricoltura per un lungo periodo di tempo. Avremo una variazione accidentale quando essa si palesa in un certo anno e dimostra di essere dovuta a cause straordinarie che si sono verificate in quell'anno e molto probabilmente non si verificheranno in avvenire.

Si capisce che la distinzione ha un'importanza radicale per il valore della ricchezza terriera, poichè le prime variazioni, a carattere permanente, incidono sul valore fondiario in quanto questo si ottiene capitalizzando la produzione, al netto delle passività, ritraibile dal fondo stesso, non in un anno ma in un gran numero di anni.

(1) L. MAROI, *op. cit.*, pag. 157.

Le altre invece, verificandosi in un anno e non ripetendosi necessariamente, riducono la produzione, cioè il reddito in quell'anno, ma non il valore fondiario.

Le prime sono dovute a fattori permanenti, come bonifiche, migliorie, trasformazioni da coltura estensiva ad intensiva, aumento di capitale investito nell'agricoltura, ecc.

Le seconde possono dipendere da fattori climatici, a volte politici. Questi ultimi possono a volte rientrare anche nella prima categoria, ma allora non operano traverso il reddito ma direttamente sul valore capitale svalutandolo, per esempio facendo temere una rivoluzione.

Si potrebbe anche, forse, identificare una variazione ciclica nella produzione, dovuta ad esaurimenti periodici della terra, ma si può ritenere che tali variazioni non influiscano sul valore fondiario, in quanto questo corrisponde alla media della produzione normale delle due fasi cicliche di fertilità e di esaurimento.

Ammesso tutto ciò occorre determinare se dal 1927 al 1928, si è verificata una variazione a carattere permanente o no. Dato che noi introduciamo nel calcolo il valore della produzione, dobbiamo considerare i fattori che lo costituiscono; precisamente: il livello dei prezzi e la quantità del raccolto.

Osserviamo il primo fattore.

I numeri indici del Consiglio Provinciale dell' Economia di Milano e del Prof. R. Bachi denotano, d'accordo, una diminuzione dal 1927 al 1928 nel livello dei prezzi attinenti all'agricoltura.

CONSIGLIO PROVINCIALE DELL'ECONOMIA DI MILANO

	DERRATE ALIMENTARI VEGETALI	PRODOTTI VEGETALI VARI
1927 (base 1923 = 100)	596.66	519.49
1928	587.65	508.17

Prof. R. BACHI

1927 (base 1923 = 100)	538.1	503.8
1928	526.6	472.0

Come si vede dalle cifre esposte la diminuzione iniziata nel 1927, perdura nel 1928.

Sarà utile analizzare anche l'andamento dei prezzi per ogni prodotto.

I cereali nel 1927 dimostrano una depressione fortissima nelle quotazioni, dovuta alla rivalutazione della lira. Nel 1928 i prezzi di questi prodotti mostrano però un rialzo, il quale è aiutato dal rialzo contemporaneo dei prezzi Nord-Americani, ma che resiste anche quando questi crollano. (1)

I prezzi del vino denotano un ribasso dovuto ad abbondante raccolto. (2)

Il prezzo dell'olio di oliva è ancora sceso, nonostante lo scarso raccolto Italiano, per opera dell'abbondantissimo raccolto internazionale. (3)

Delle colture industriali, si può avere qualche notizia sulla canapa. Per questa si osserva nel 1928 una certa sostenutezza nei prezzi, dovuta anche alla previsione di uno scarso raccolto. (4)

Ciò che qui importa notare è che l'andamento dei prezzi non è dipeso da fattori permanenti, come è successo nel 1927, ma in gran parte da fattori accidentali, come scarsi raccolti o abbondanti, in casa nostra od altrove.

Da questo fattore, livello dei prezzi, non dobbiamo quindi aspettarci un'opera di trasformazione nel valore della produzione se non assai lieve, cioè per quella parte che si ricollega al generale discendere dei prezzi voluto dalla nostra politica, che si prolunga nel 1928.

Vediamo l'altro fattore: la quantità.

Si può osservare che, a parità di altre condizioni, esso è reciproco del primo, in quanto i prezzi siano comandati unicamente dall'abbondanza o scarsità del raccolto di casa, ma non verificandosi questa legge è chiaro che i due fattori possono agire indipendentemente.

Sarà utile dare uno sguardo all'andamento della nostra produzione nel 1928, espressa in quantità.

(1) Banca Commerciale Italiana. *Movimento economico dell'Italia — Dati statistici per il 1928*, Milano 1929, pag. 149.

(2) Id. id., pag. 159.

(3) Id. id., pag. 167.

(4) Id. id., pag. 174.

Risultati dell'annata agricola 1927-28 (1)

(*Migliaia di quintali*)

	1923-27	1927	1928
Cereali	98.951	90.470	96.125
Piante industriali	26.046	21.381	29.873
Patate	20.336	19.453	14.898
Legumi	5.791	5.138	6.485
Ortaggi	15.302	13.625	(2) 13.500
Vino (3)	43.351	35.650	46.823
Olio di oliva	1.855	1.602	(2) 2.400
Frutta e agrumi	17.092	16.850	17.428
Castagne	4.790	4.280	5.579
Foraggi	229.100	217.200	202.700

(1) Banca Commerciale Italiana, *op. cit.*, pag. 110.
(2) Previsioni.
(3) Migliaia di lire.

Tirando le somme dei dati esposti, vediamo che dal 1927 al 1928 si nota un aumento quasi generale. Ma confrontando le cifre del 1928, con quelle medie del periodo 1923-27 notiamo che le prime rimangono quasi sempre inferiori alle seconde. Ciò significa che la diminuzione verificatasi nel 1927, si va risolvendo, ma ancora nel 1928 se ne trovano tracce.

Risulta anche da questa comparazione, come dalla impressione della nostra crescente attività agricola, che le minori cifre del 1928, in confronto di quelle medie del periodo 1923-27, non dipendono da fattori permanenti ma da fattori accidentali, la cui influenza si fa sentire transitoriamente.

Considerando il nostro calcolo sotto questo unico aspetto, si deduce che adottando come produzione, da cui ricavare il reddito

da capitalizzare, quella del 1928, si ottiene una valutazione inferiore alla realtà, tanto più inferiore se si parte da quella del 1927. Però non bisogna dimenticare l'altro fattore: i prezzi.

Questi dal 1927 al 1928 sono discesi ancora, in parte per cause accidentali, ma in parte anche per quella generale tendenza al ribasso che probabilmente continuerà.

Concludendo noi crediamo che adottando la valutazione del FRANCIOSA, riferentesi al 1927, l'errore a cui andremo incontro, sarà certamente lievissimo, poichè la discesa dei prezzi compensa il maggior raccolto del 1928; in più se questo è sempre inferiore al normale, anche questa circostanza sarà compensata dalla tendenza che hanno i prezzi a scendere ancora, che risulta dai numeri indici del 1929.

Calcolata così la produzione occorre, seguendone il processo di distribuzione, passare al beneficio fondiario.

A questo scopo, seguendo la terminologia del TASSINARI, dobbiamo identificare il reddito fondiario, e depurarlo dall'eventuale profitto.

Il SACERDOTE, in un recentissimo calcolo della ricchezza italiana, (1) riduce la produzione lorda, la cui valutazione egli prende dal PORRI, ai $^3/_5$ che ritiene rappresentino il reddito netto, questo poi lo capitalizza con un saggio di capitalizzazione uguale a sette. Come nota il MAROI (2), il SACERDOTE ottiene il reddito netto detraendo dal lordo unicamente la quota di reintegrazione dei capitali impiegati, il che non costituisce una riduzione completa, in più non si capisce cosa rappresenti « prendere sette volte il reddito netto ». È preferibile seguire una via più chiaramene logica, cioè: stabilire il beneficio fondiario e capitalizzarlo.

Dobbiamo notare subito che dal 1908 ad ora la coltura si è certamente intensificata, quindi, in base al principio della diminuzione della proporzione del reddito netto con l'intensificarsi della coltura, dobbiamo ammettere a priori che la percentuale che rappresenta il reddito netto rispetto alla produzione lorda, sarà minore nel 1928 di quello che era nel 1908.

Il Prof. C. GINI, appunto in quell'epoca fissava a 25 % (3) tale percentuale. Considerando questa proporzione per allora troppo bassa,

(1) SACERDOTE ALBERTO. *Ricchezza e reddito privati in Italia*, Pinerolo, 1928, pag. 33.

(2) L. MAROI, *op. cit.*, pag. 47.

(3) C. GINI. *L'ammontare e la composizione* ecc., pag. 173.

il MAROI propone, per gli anni in corso il 24 %. (1) Noi, attenendoci invece all'affermazione del primo autore, crediamo attendibile una percentuale del 23 %, al netto delle imposte. I risultati dell'indagine eseguita dal TASSINARI su alcuni tipi di aziende, confermano, del resto, questa percentuale, se si tiene conto del fatto che quella a cui giunge il TASSINARI rappresenta il reddito fondiario, cioè beneficio più profitto e quindi va ridotta ancora.

In base a questa proporzione il beneficio fondiario viene ad essere di 9.775 milioni.

A confermare questa cifra possiamo tentare un secondo calcolo. Dalla revisione catastale, ordinata con legge del Giugno 1923, è risultato che la rendita censuaria imponibile, in moneta aurea è di lire 1.466.912.000. (2)

In moneta attuale tenendo conto della variazione del potere d'acquisto della moneta, abbiamo che tale rendita censuaria corrisponde ad una cifra di 7.202.538.000 lire. (3)

Questa cifra rappresenta però soltanto il reddito dominicale accertato agli effetti delle imposte, il quale è ben lontano dal rappresentare l'effettivo. In altre parole dobbiamo per trovare questo ultimo, determinare un coefficiente d'evasione. Per il 1908, il Prof. C. GINI ammetteva una evasione minima del 25 %, (4) tenendo conto da una parte dell'inasprimento delle imposte, che tende ad accentuare l'evasione, dall'altra del maggior spirito di disciplina che è penetrato negli Italiani, e della maggior cura degli agenti fiscali, crediamo che si possa ammettere questo coefficiente di evasione anche per il nostro calcolo. In base a questo, dunque, il reddito effettivo sarebbe di 9.003.172.000.

Se si considera che l'evasione del 25 % è un minimo (5), come si può dedurre da molte osservazioni fatte in appositi studi, si capisce come la cifra così trovata confermi l'altra a cui ci atterremo nel nostro calcolo.

(1) L. MAROI, *op. cit.*, pag. 46.

(2) MINISTERO DELLE FINANZE. *La gestione delle imposte dirette dal* 1914 *al* 1925, pag. 88-89.

(3) La cifra è ottenuta tenendo conto del livello dei prezzi del 1928 in confronto dell'anteguerra, secondo il numero indice del Consiglio Provinciale dell'Economia di Milano.

(4) C. GINI. *L'ammontare e la composizione* ecc., pag. 150.

(5) C. GINI. *L'ammontare e la composizione* ecc., pag. 216-217. L'A. afferma che l'evasione della tassa di successione dei terreni non è molto inferiore al 50 %. Questa evasione fa capire come anche l'altra per l'imposta debba essere certamente superiore al 25 %.

In base a queste considerazioni, dunque, riteniamo di poter ammettere la cifra di 9.775 milioni come rappresentativa del beneficio fondiario.

Rimane sempre da determinare l'ultimo elemento del calcolo, il saggio di capitalizzazione.

Il Prof. C. GINI, nel 1908, ammetteva un saggio di capitalizzazione del 100 per 4, e basava questa sua affermazione sul fatto che la rendita pubblica dava un interesse del 3.5 %. (1)

Attualmente la rendita pubblica dà il 5 %, ma le condizioni di fiducia nei titoli della rendita stessa sono peggiorate dall'anteguerra, quindi si capisce come, dipendendo il saggio ora detto dalla minor fiducia, si debba fissare quello per il rendimento delle terre ad un livello non superiore. Così il MAROI accetta un saggio corrispondente al 5 % di interesse. (2)

Il saggio del reddito fondiario, risultante dalle indagini del TASSINARI, non si discosta, anch'esso, da tale cifra. (3)

Crediamo quindi opportuno capitalizzare il beneficio fondiario trovato al 100 per 5.

Eseguendo il calcolo il valore fondiario risulta di 195,5 miliardi, da cui detraendo 27 miliardi di bestiame (4) rimangono 168,5.

(1) C. GINI, *L'ammontare e la composizione* ecc., pag. 173.

(2) L. MAROI, *op. cit.*, pag. 46.

(3) Crediamo opportuno presentare i saggi del reddito, che abbiamo calcolato come rapporto fra il valore fondiario e il reddito fondiario, in base alle cifre fornite dal TASSINARI, nel suo studio più volte citato per il 1922, pagine 151-156.

Tipi di aziende irrigue lombarde-piemontesi	6,54 %
	6,39 %
Tipi di aziende asciutte emiliane con colture industriali	7,90 %
	8,46 %
Tipi di aziende a coltura promiscua, Italia centrale	6,82 %
	10,08 %
	4,86 %
	3,70 %
Tipi di aziende a coltura legnosa specializzata	8,87 %
	6,56 %
	2,34 %
	6,69 %
Tipi di aziende latifondistiche a prevalente cerealicoltura	5,31 %
	4,90 %
Tipi di aziende latifondistiche cerealicole-pastorali	5,32 %

(4) Vedi appresso, la valutazione del bestiame.

Ma questa cifra non rappresenta ancora la proprietà fondiaria in mano di privati; per ottenere questa dobbiamo detrarre il valore di quella appartenente a persone giuridiche. Accetteremo per questa detrazione la proporzione fissata dal prof. C. Gini, in base ai dati della Direzione Generale delle imposte dirette, cioè ammetteremo che il 92 % della proprietà totale sia dei privati. (1)

Anche per questa via risulta, dunque, che la ricchezza terriera dei privati si aggira verso i 155 miliardi.

3). — Controllo ai calcoli precedenti in base ad un'inchiesta presso le Cattedre ambulanti di agricoltura

Infine, come ultimo controllo ai nostri calcoli, abbiamo fatto una piccola inchiesta presso le cattedre ambulanti di agricoltura, allo scopo di conoscere il prezzo venale medio per ettaro risultante ai Professori Direttori delle cattedre stesse, e il saggio medio di capitalizzazione da usarsi per passare dal reddito netto al valore delle terre italiane.

I dati che sono stati chiesti corrispondono alla seguente ripartizione delle terre:

Seminativi;
Vigneti;
Oliveti;
Agrumeti;
Frutteti, Gelseti, Mandorleti, Vivaieti;
Boschi (esclusi i castagneti);
Castagneti;
Prati stabili;
Incolto produttivo.

Non tutte le cattedre hanno risposto alla nostra circolare in tempo utile, ma il numero di quelle che hanno risposto è sempre soddisfacente. Ai Direttori di queste cattedre, approfittando del momento, indirizziamo un ringraziamento cordiale per l'aiuto prestatoci.

In base ai dati fornitici, per quelle provincie la cui cattedra ha risposto, abbiamo calcolato come media ponderata, in cui i pesi sono costituiti dalla superficie adibita ad ogni coltivazione, i prezzi medi venali per ettaro e i saggi di capitalizzazione medi.

(1) C. Gini. *L'ammontare e la composizione* ecc., pag. 174.

Non è stato però possibile da questa media provinciale passare ad una media generale del Regno essendo rischioso estendere i dati limitati ad una zona ad altre anche vicine.

Ad ogni modo i dati che seguono, distaccandosi poco, relativamente, da quelli da noi stabiliti per altra via per l'intiero Regno, confortano l'attendibilità di questi.

Per comodità di confronto abbiamo raggruppato le provincie per regioni e secondo la più ampia ripartizione fra Italia Settentrionale, Centrale, Meridionale, Insulare.

VALORE VENALE MEDIO PER ETTARO
SAGGIO DI CAPITALIZZAZIONE MEDIO

PROVINCIE	VALORE VENALE MEDIO (per ettaro) lire	SAGGIO DI CAPITALIZZAZIONE medio
ITALIA SETTENTRIONALE		
PIEMONTE:		
Provincia di Cuneo	14.048	4,35 %
Provincia di Torino	11.532	3.43 %
LIGURIA:		
Provincia di Genova	5.243	3 — %
LOMBARDIA:		
Provincia di Como	8.798	4 — %
Provincia di Pavia	6.676	3,82 %
Provincia di Cremona	8.918	5 — %
Provincia di Milano	17.986	4,96 %
Provincia di Brescia	11.795	5,82 %
VENETO:		
Provincia di Vicenza	12.238	5 — %
Provincia di Rovigo	10.335	4,14 %
VENEZIA TRIDENTINA:		
Bolzano	7.177	2,55 %
EMILIA:		
Provincia di Piacenza	12.777	7 — %
Provincia di Ferrara	6.637	5,88 %

PROVINCIE	VALORE VENALE MEDIO (per ettaro) lire	SAGGIO DI CAPITALIZZAZIONE medio
ITALIA CENTRALE		
MARCHE:		
Ancona	9.000	6 — %
UMBRIA:		
Terni	2.739	5,02 %
TOSCANA:		
Provincia di Arezzo	3.493	5 — %
Provincia di Grosseto	2.290	5,49 %
Provincia di Lucca	11.315	4,04 %
Provincia di Pistoia	..	5 — %
ITALIA MERIDIONALE		
CAMPANIA:		
Provincia di Benevento	2.810	3,29 %
PUGLIE:		
Provincia di Taranto	4.317	7 — %
CALABRIE:		
Provincia di Cosenza	4.765	6.29 %
ITALIA INSULARE		
SARDEGNA:		
Provincia di Cagliari	10.374	7,96 %
Provincia dl Sassari	2.433	6,4 %
SICILIA:		
Provincia di Siracusa	7.832	7,38 %
Provincia di Agrigento	4.385	5,05 %
Provincia di Trapani	12.924	3,80 %
Provincia di Caltanissetta	4.714	5 — %

IV.

VALORE DELLE MINIERE E TORBIERE

Per questa categoria di ricchezza essendosi verificata da una parte una diminuzione dei prezzi, dall'altra un aumento di produzione, possiamo ritenere inalterata la valutazione in 5 miliardi data dal Prof. C. Gini (1) per il 1925.

V.

VALORE DEI FABBRICATI URBANI DI PROPRIETÀ PRIVATA

Perdurano le difficoltà di calcolare il valore dei fabbricati urbani, che si trovava contro il Prof. Gini, nel suo calcolo della ricchezza italiana nel 1925, date principalmente dalla limitazione degli affitti che pone in un equilibrio artificiale il valore venale delle case.

Il risultato dei nostri calcoli avrà quindi un carattere d'incertezza inevitabile.

1°) Cominceremo con il valutare il valore dei fabbricati urbani basandoci sul reddito imponibile accertato agli effetti dell'imposta sui fabbricati; in seguito, dato che la legge concede molte esenzioni, sarà necessario integrare il calcolo con una valutazione diretta dei fabbricati esenti dall'imposta.

Riguardo alla prima parte del calcolo dobbiamo tenere presente diverse circostanze che obbligano a portare alcune correzioni alla cifra che rappresenta il reddito imponibile prima di fare la capitalizzazione.

Innanzitutto, il reddito imponibile rappresenta i tre quarti del reddito accertato, defalcando l'Amministrazione delle Imposte appunto ¼ da questo ultimo per tener conto delle spese di manutenzione dei fabbricati. (2)

In secondo luogo il reddito accertato non rappresenta il reddito effettivo, a causa dell'evasione. Il Prof. C. Gini (3) ammette che il reddito accertato sia i $^3/_4$ dell'effettivo nel 1908; noi, considerando che

(1) C. Gini. *A comparison of the wealth* ecc., pag. 6.
(2) C. Gini. *L'ammontare e la composizione* ecc., pag. 174.
(3) Id. id., pag. 175.

la facilità di evasione in questi ultimi anni è assai diminuita, potremo supporre la frazione di reddito che sfugge uguale a $\frac{1}{4}$ del reddito accertato, vale a dire tale da compensare le spese di manutenzione.

Infine dobbiamo tener presente che una parte del reddito dei fabbricati non appartiene a privati; in base ai dati forniti dal Prof. C. GINI (1) porremo questa frazione uguale al 14 %.

Prima di passare al calcolo è bene osservare che non si corrono rischi di ripetere beni già contati o che lo saranno in altre voci, poichè i fabbricati rurali, essendo esenti dalle imposte sui fabbricati, non figurano nella presente valutazione come sorgenti di reddito, inoltre i fabbricati industriali (a partire dal 30 dicembre 1923) non sono più soggetti all'imposta sui fabbricati, ma considerati fra le fonti di reddito colpite dall'imposta di ricchezza mobile.

Il reddito accertato che viene qui assunto come base del calcolo, è quello risultato dalla revisione e rivalutazione generale provvisoria dei redditi sui fabbricati (2) ordinata con decreto 30 dicembre 1923.

I criterî che hanno guidato questa revisione sono i seguenti:

PER OGNI 100 LIRE DI REDDITO IMPONIBILE ACCERTATO NEL PERIODO	IL NUOVO REDDITO IMPONIBILE PROVVISORIO FU DETERMINATO IN LIRE
1° gennaio 1891-31 dicembre 1910	400
1° gennaio 1910-31 dicembre 1918	350
Anni 1919 e 1920	250
Anni 1921 e 1922	150
Dal 1° gennaio 1923 in poi	100

La scala discendente fu fatta per tenere conto della circostanza che la moneta si era andata man mano svalutando. (3)

Il reddito totale accertato imponibile che noi prenderemo è l'ultimo pubblicato cioè quello riferentesi al 1926, valutato in 2.955.566.000

Secondo quanto è stato esposto esso rappresenta un reddito effettivo di 3.694.457.500.

Prima di procedere nel calcolo è necessario però detrarre da tale cifra il carico delle imposte e sovraimposte.

(1) C. GINI. *L'ammontare e la composizione* ecc., pag. 174.

(2) L'influenza della revisione andata in vigore nel 1925 è visibilissima se si considerano i redditi imponibili accertati di anno in anno: il valore del 1925 mostra un aumento enorme in confronto di quelli precedenti. L'aumento continua ma lieve nel 1926 mostrando semplicemente che è aumentato il numero degli articoli di ruolo.

(3) L. EINAUDI. *Scienza delle Finanze*, pag. 299.

Nel 1926 esse ammontarono in totale a 770.598.000 (1), cioè il reddito netto risulta di 2.913.859.500, da cui detraendo il 14 %, come non appartenente a privati, abbiamo 2.505.918.710.

Si tratta ora di determinare un giusto coefficiente di capitalizzazione.

Essendo quello dei terreni al 100 per 5, crediamo poter fissare questo al 100 per 6. (2)

Adoprando questo coefficiente otteniamo un valore di circa 41.598 milioni.

Bisogna ora valutare i fabbricati esenti.

Tali esenzioni sono a favore di ricostruzioni di zone terremotate, di zone danneggiate dalla guerra, a favore di case popolari economiche, infine a favore delle case costruite dopo il 5 luglio 1919. (3)

Il calcolo dei fabbricati compresi in queste esenzioni è terribilmente difficile poichè mancano del tutto dati statistici su cui basarsi.

Si deve necessariamente ricorrere a delle presunzioni.

Il Prof. C. GINI valutava il valore dei fabbricati nel 1908 in 16 miliardi, (4) nel 1914 in 20, e nel 1917 in 21. (5)

Egli ammetteva cioè dal 1908 al 1914 un incremento annuo di 660 milioni, dal 1914 al 1917 un incremento di 330 milioni.

È necessario giudicare se è possibile calcolare l'incremento edilizio dal 1919, dall'anno, cioè, in cui sono incominciate le esenzioni, uguale a uno di questi due incrementi oppure ad uno da essi dedotto.

Per oggettivare l'impressione che ognuno può avere sull'attività edilizia del dopo guerra in confronto dell'anteguerra abbiamo calcolato un indice dell'attività edilizia delle grandi città, che si può supporre significativo per l'intera nazione.

(1) « Annuario Statistico Italiano », 1928, pag. 337.

(2) Il Prof. GINI nella sua valutazione della ricchezza privata italiana del 1908, ammetteva che fra il saggio del reddito dei terreni e quello dei fabbricati ci fosse una proporzione differente da quella ammessa in questo calcolo: egli fissava quello dei terreni al 4 %, e quello dei fabbricati a 5,4 %. Seguendo la stessa proporzione avremmo dovuto prendere un saggio superiore al 6 %, ma ci siamo attenuti a questo considerando che i terreni non si trovano sottoposti a quelle limitazioni negli affitti a cui invece si trovano sottoposti i fabbricati.

(3) MINISTERO DELLE FINANZE. *La gestione delle imposte dirette dal 1914 al 1925*. Decreto 8 marzo 1923, n. 695.

(4) C. GINI. *L'ammontare e la composizione* ecc., pag. 176.

(5) C. GINI. *Problemi sociologici della guerra*, pag. 264.

Questo indice rappresenta il numero complessivo delle stanze costruite nei vari anni, nelle seguenti grandi città: Torino, Milano, Venezia, Genova, Firenze, Roma, Napoli, essendo fatto uguale a 100 il numero delle camere costruite nel 1913. (1)

NUMERI INDICI DELL'ATTIVITÀ EDILIZIA IN ALCUNE GRANDI CITTÀ

1904	90.4	1919	40.6
1906	55.6	1920	74.4
1908	72.6	1921	54.1
1909	121.4	1922	96.5
1913	*100.0*	1923	152.3
1914	118.2	1924	195.0
1915	56.8	1925	247.2
1916	35.2	1926	272.5
1917	21.5	1927	258.9
1918	13.3		

Dal 1908 al 1914 l'indice medio è stato 98; dal 1914 al 1917, 70; dal 1919 al 1927, 143.4; ricordando gli incrementi accettati dal Prof. C. GINI per i primi due periodi si può ammettere, in base agli indici calcolati, ai prezzi dell'anteguerra, un incremento medio annuo dal 1919 in poi di 800 milioni, in totale, cioè, di 7.200 milioni (2). Ai prezzi del materiale da costruzione nel dopo guerra (considerando che l'indice dei prezzi del materiale da costruzione è nel 1928=521.30) (3) si otterrebbero 37.512 milioni, che sommati ai 42 miliardi già calcolati ci danno in cifra tonda 80 miliardi.

2º) È possibile tentare di valutare i fabbricati urbani in una seconda maniera.

(1) UGO GIUSTI. *Abitazioni ed attività edilizia nelle grandi città italiane prima, durante e dopo la guerra*, « Economia », vol. II, n. 12, anno VI.

L'A. raccoglie i dati dagli annuari statistici delle varie città, sull'attività edilizia, rappresentata dal numero di stanze dichiarate abitabili ogni anno.

In una grossolana approssimazione si può considerare che l'attività edilizia in alcune grandi città sia significativa per l'attività edilizia di tutto il Regno.

A partire dal 1919 incluso, l'indice da noi costruito comprende anche il numero di stanze costruite in Trieste, che invece era stato detratto per gli anni precedenti.

Ciò è stato fatto per tener conto dell'apporto delle terre redente all'attività edilizia.

(2) Tale incremento è stato da noi calcolato in base ai valori dei fabbricati risultanti dai calcoli del Prof. C. GINI per gli anni 1908, 1914, 1917.

(3) CONSIGLIO PROVINCIALE DELL'ECONOMIA DI MILANO.

Naturalmente, essendo questi calcoli assai incerti, l'uguaglianza dei risultati ottenuti per vie differenti e il numero delle vie seguite sono elementi essenziali per potere giudicare dell'attendibilità dei risultati stessi.

Questa seconda via di calcolo si basa sull'ipotesi che il valore dei fabbricati sia variato dal 1914 al 1928 secondo un coefficiente che tiene conto dell'aumento dei prezzi e dell'aumento numerico dei fabbricati.

Tale coefficiente composto è costituito dal prodotto di due coefficienti semplici di cui uno tiene conto del primo fattore di variazione e l'altro del secondo.

La determinazione del primo coefficiente non presenta difficoltà: lo possiamo supporre uguale al rapporto fra il livello del fitto medio pagato da una famiglia nel 1914 a quello pagato nel 1928.

Così facendo veniamo a supporre che il valore degli stabili sia variato come i fitti che da essi si possono ricavare: tale ipotesi è senza dubbio accettabile.

Il secondo è più difficile a determinarsi. Noi lo abbiamo fissato uguale al rapporto fra il numero totale delle famiglie esistenti nel 1928 e quello delle famiglie esistenti nel 1914. La correlazione fra numero di famiglie e l'ammontare dei fabbricati è accettabile facilmente. Ciò che può essere dubbio è se è accettabile di assumere per il calcolo il numero totale delle famiglie.

Si può infatti osservare che nel numero totale di famiglie è computato anche quello delle famiglie abitanti gli edifici adibiti ai fondi rustici e quindi già compresi nella stima dei terreni. D'altra parte non è possibile valutare l'ammontare di queste famiglie poichè i censimenti danno soltanto il numero delle famiglie dedite all'agricoltura indipendentemente dal fatto che abitino o no il fondo che coltivano, mentre, come è noto, molte famiglie di contadini, sopratutto nell'Italia meridionale, non abitano i fondi, ma in case di città o di villaggi.

Tutto ciò però non può arrestare il nostro calcolo. Infatti supporre che il coefficiente d'aumento numerico dei fabbricati dal 1914 al 1928 sia uguale al rapporto fra il numero totale delle famiglie nel 1914 e quello delle 1928, significa supporre che l'aumento delle famiglie non abitanti i fondi rustici sia avvenuto nella stessa proporzione in cui è avvenuto quello delle famiglie abitanti i fondi rustici. Tale ipotesi non è esatta poichè dall'anteguerra al dopo guerra la concentrazione della popolazione è andata aumentando a scapito delle

campagne, ma noi crediamo che l'approssimazione che offre sia sufficiente ai nostri scopi. (1) Naturalmente il risultato verrà approssimato per difetto.

Premesso tutto ciò possiamo passare al calcolo. Basandoci sulle variazioni dell'indice generale dei prezzi e sulla percentuale che rappresentava prima della guerra e dopo, sul reddito medio familiare, la somma pagata per l'affitto, possiamo giungere a calcolare il primo coefficiente che cerchiamo.

Avanti la guerra la percentuale del fitto sul reddito medio era di 11.4 %, nel dopo guerra di 4.94 %. (2) Cioè la spesa costituita dal fitto nel dopo guerra ha perso metà circa dell'importanza che aveva nell'anteguerra.

D'altra parte possiamo supporre che il reddito familiare sia aumentato come il livello generale dei prezzi, in media, cioè di 4,91. (3)

Allora in formule, chiamando F, il fitto d'anteguerra, F' quello di dopoguerra, R il reddito d'anteguerra abbiamo che:

$$F = \frac{11{,}4}{100} R \quad \text{e} \quad F' = \frac{4{,}94}{100} R.\ 4{,}91;$$

il coefficiente che cerchiamo sarà allora:

$$J = \frac{F'}{F} = \frac{4{,}94 \cdot 4{,}91}{11{,}4} = 2{,}12$$

Possiamo fare anche un altro ragionamento, il reddito medio per articolo di ruolo, accertato era nel 1914 di 215 lire, nel 1926 era di 791 lire: il rapporto sarebbe di 3.67. (4)

Dobbiamo però osservare che l'aumento, in base al reddito medio per articolo di ruolo, risulta superiore al precedente anche perchè nel dopoguerra l'accertamento è stato più accurato e l'evasione minore. Quindi essendosi nel valore dei fabbricati per il 1914 già tenuto conto dell'evasione, non è rigoroso usare per il calcolo dell'aumento di questo valore il coefficiente di 3,67, ora trovato.

(1) MARIO SAIBANTE. *La concentrazione della popolazione.* « Metron », vol. VII, n. 2, 1928.

(2) « La proprietà edilizia italiana ». Rivista mensile della Federazione Fascista della proprietà edilizia. Gennaio-Febbraio 1929; a pag. 6, si danno, in un articolo del Commissario straordinario ENRICO PARISI, gli indici da noi adoprati.

(3) Livello generale dei prezzi, media per il 1928. Costruito dal CONSIGLIO PROVINCIALE DELL'ECONOMIA DI MILANO.

(4) MINISTERO DELLE FINANZE. *La gestione delle imposte dirette dal 1914 al 1925.* « Annuario Statistico Italiano », *1928.*

Considerando l'incertezza dei due coefficienti appare prudente scegliere un valore intermedio, ma molto più vicino al secondo che al primo. Assumeremo quindi per il nostro calcolo il valore di 3,50. Procediamo ora al calcolo del secondo coefficiente che deve tener conto dell'aumento numerico dei fabbricati.

Esso viene ad essere uguale al rapporto fra il numero delle famiglie esistenti in Italia nel 1928 e quello per il 1914, e risulta di 1,15. (1)

Il coefficiente composto dei due separatamente calcolati risulta finalmente di 4,025.

Partendo dalla valutazione dei fabbricati in 20 miliardi fatta per il 1914 dal prof. C. GINI (2) otteniamo per il 1928 la cifra di 80,5, miliardi, in cifra tonda 80 miliardi.

3°) Possiamo infine tentare un'ultima via di valutazione della ricchezza edilizia. Basandoci sulla valutazione dei fabbricati nel 1925 eseguita dal Prof. C. GINI, possiamo passare al valore attuale tenendo conto dei prezzi e dell'incremento edilizio dal 1925 al 1927 compreso. La cifra data dal Prof. C. GINI è di 90 miliardi. (3) Si vede immediatamente che con questa si accorda quella di 80 precedentemente da noi trovata, essendosi verificata negli ultimi tre anni una diminuzione dei prezzi e un netto incremento di costruzioni.

Possiamo infatti supporre che il valore delle case sia variato come il livello dei prezzi del materiale da costruzioni, cioè come da 655,21 nel 1925 a 521.30 nel 1928. (4)

Eseguendo la proporzione si trova che il valore delle case che nel 1925 era sui 90 miliardi nel 1928 si può considerare sui 71 miliardi e mezzo (71.6). Aggiungendo a questi il valore dei nuovi fabbricati si giunge facilmente alla cifra di 80 miliardi da noi trovata, anzi si supera giungendo agli 83 miliardi.

4°) La concordanza dei risultati ottenuti per tre vie diverse, costituisce una conferma se non dell'esattezza del risultato, almeno della approssimazione soddisfacente che questo ha raggiunto.

(1) Nel 1914 esistevano in Italia 7.539.646 famiglie (*Censimento del 1911*, vol. VI, pag. 448) nel 1921 8.594.223 (*Censimento del 1921*, Relazione generale, pag. 74). Per calcolare il numero delle famiglie esistenti nel 1928 ci siamo basati sul numero medio di componenti una famiglia nel 1921, e sull'incremento della popolazione dal 1921 al dicembre 1927 (« Bollettino di Statistica », Marzo, 1929) il numero trovato è di 8.696.494.

(2) C. GINI. *Problemi sociologici della guerra*, pag. 264.

(3) C. GINI. *A comparison of the wealth* ecc., pag. 6.

(4) Indici dei prezzi calcolati dal CONSIGLIO PROVINCIALE DELL'ECONOMIA DI MILANO.

Dovendo giudicare del senso dell'errore, saremmo portati a credere che esso sia negativo cioè che il valore trovato rimanga ancora inferiore alla realtà. Questa convinzione è basata su alcune considerazioni che vengono spontanee nell'eseguire il calcolo stesso.

Si può osservare, per esempio, che la qualità delle case è migliorata.

Tutte le nuove costruzioni, lo può testimoniare l'esperienza quotidiana di ognuno, presentano dei caratteri di igiene, di eleganza, superiori a quelli che presentavano le case costruite nell'anteguerra, dimodochè a parità di dimensioni, esse rappresentano un maggior valore. D'altra parte noi, per calcolare l'incremento edilizio, ci siamo basati unicamente sull'elemento dimensioni, cioè numero delle stanze.

Si può osservare, è vero, che essendo le case del dopo guerra, specialmente quelle di tipo economico, di grande mole, in esse le spese generali costituiscono un minore aggravio per stanza, essendo ripartite su di un numero maggiore di stanze, ma è anche da ritenersi questa influenza inferiore di gran lunga all'altra per cui quasi tutte le nuove case hanno riscaldamento e bagno, e sono più confortabili.

Scarsa invece deve essere l'influenza della variazione dell'affollamento di abitanti nello stesso appartamento. Alcuni dati raccolti dal Giusti ci danno appunto questa impressione. (1)

Se così non fosse, è evidente che il calcolo basato sul valore medio per famiglia perderebbe completamente di significato.

Queste considerazioni, dànno appunto la sensazione che la valutazione in 80 miliardi rimanga inferiore alla realtà.

Può essere utile aggiungere, ora, qualche parola, per spiegare il contrasto fra la presente valutazione e quella del Mortara per il 1927 in 50 miliardi (2)

Evidentemente è difficile identificare cause che possano aver fatto variare così nettamente il valore dal 1927 al 1928: l'incremento edilizio del 1927 può essere compensato dalla diminuzione dei prezzi, e d'altra parte non potrebbe essere tale da giustificare una differenza di diecine di miliardi. È da ritenersi quindi che tale differenza sia dovuta a diversità di metodi di calcolo, e non a circostanze che abbiano fatto variare il valore stesso.

(1) U. Giusti. Articolo citato. Egli dice: « Preso così all'ingrosso l'affollamento medio trovato nel 1921 nelle abitazioni fino a 5 stanze non differisce sensibilmente da quello verificato nel 1911 », pag. 507.

(2) G. Mortara. *La ricchezza nazionale, il reddito,* ecc.

Alcune considerazioni mostreranno però, che la cifra del MORTARA rimane certamente inferiore al vero, in forte misura. Egli nello stesso calcolo della ricchezza dà per i terreni 160 miliardi come valutazione nel 1927. (1)

Ricordando le valutazioni del Prof. C. GINI, che sono le più rigorose ed attendibili, si nota come il valore dei fabbricati rappresenti un po' meno della metà del valore dei terreni in mano dei privati, un po' più di un terzo dei terreni in totale. (2)

Ora il rapporto fra valore dei fabbricati e terreni non varia a capriccio, ma segue una linea di sviluppo fissata dall'andamento delle condizioni demografiche, sociali ed economiche della nazione.

Queste, dato l'aumento della popolazione, non possono che aver agito uniformemente dal punto di vista della rendita, tanto sui terreni, quanto sui fabbricati, ma solo sui fabbricati possono aver determinato un notevole aumento di quantità.

Si capisce quindi come naturalmente la tendenza nei movimenti dei due valori sarebbe stata diretta a diminuire la differenza fra valore dei terreni e quello dei fabbricati.

Considerando quelli dati dal MORTARA si vede invece che i fabbricati rappresentano un po' meno di un terzo del valore dei terreni. Si può dire che le limitazioni degli affitti, hanno impedito l'aumento naturale del valore dei fabbricati, ma allora bisogna anche osservare che la maggior frazione che va ai lavoratori del prodotto dell'agricoltura, in altre parole l'aumento del costo della mano d'opera, più sensibile nell'agricoltura che nella manutenzione dei fabbricati, può aver controbilanciato, almeno in parte, questa circostanza sfavorevole ai fabbricati. Si osserva che le imposte si sono aggravate di più sui fabbricati che sui terreni, (3) e ciò è indubbiamente vero, ma l'influenza di tale circostanza potrà aver tenuto costante il rapporto fra valore dei fabbricati e dei terreni, che per le altre avrebbe teso ad aumentare, ma non sembra sufficiente a spiegare una diminuzione nettissima come quella che apparirebbe dai calcoli del MORTARA.

La nostra valutazione non altera quasi il rapporto fra valore dei dei fabbricati e terreni, e appunto per questo sembra più esatta.

(1) G. MORTARA. *Op. cit.*, pag. 397.

È necessario aggiungere che il MORTARA valuta la proprietà terriera nazionale e non solo la privata.

(2) Nel 1908 il valore dei terreni era in totale 40, e quelli dei privati 37 miliardi, il valore dei fabbricati 16.

Nel 1914 il valore dei terreni dei privati era 44, il valore dei fabbricati 20.

(3) L. EINAUDI. *Scienza delle Finanze*, pag. 299.

VI.

VALORE DEL BESTIAME

Mancano assolutamente dati numerici sul bestiame, per epoche recenti. L'ultimo censimento si riferisce al 1918, i calcoli specifici eseguiti su questo argomento sono pure ormai antiquati; per esempio il calcolo eseguito dal FOTTICCHIA, uno dei più accurati, si riferisce al 1921-22, (1) quello del NICCOLI ugualmente al 1921-22. (2) Inoltre tali calcoli valutano la produzione zootecnica e non il patrimonio italiano costituito dal bestiame.

Siamo stati quindi costretti a basarci su calcoli passati, cercando di determinare, congetturalmente l'incremento.

Il Prof. C. GINI valutava nel 1925 il bestiame a 35 miliardi; (3) supponendo i prezzi invariati e l'incremento annuo uguale a quello ammesso dallo stesso autore fra il 1908 e il 1914, (4) si giunge per il 1928 alla cifra di 36 miliardi.

Tale cifra va naturalmente ridotta, tenendo conto della diminuzione dei prezzi. Come indice di questi possiamo assumere l'indice dei prezzi delle derrate alimentari animali calcolato dal Consiglio Provinciale dell'Economia di Milano.

La somma di 36 miliardi va allora ridotta in proporzione alla diminuzione degli indici di tali prezzi. Eseguendo il calcolo si raggiunge la cifra di 27 miliardi.

Data l'incertezza del precedente calcolo è utile eseguire una riprova: la eseguiremo supponendo che fra la produzione zootecnica e il valore del bestiame si verifichi nel 1928 lo stesso rapporto che si verificava per il 1914.

Per tale anno il FOTTICCHIA valutava la produzione zootecnica a 2.339 milioni, (5) d'altra parte il Prof. C. GINI valutava il bestiame a 4.890 milioni. (6) Il rapporto è di 2.09.

(1) G. FOTTICCHIA. « Annali del Ministero dell'Agricoltura », anno II, n. 2, 30 giugno 1922.

(2) V. NICCOLI. *Economia rurale. Agrotimesia e computisteria agraria*, 2ª edizione curata e aggiornata dal Dott. NELLO NICCOLI, Torino, 1927.

(3) C. GINI. A *comparison of the wealth* ecc., pag. 6.

(4) C. GINI. *Problemi sociologici della guerra*, pag. 253. L'A., in base alle notizie raccolte dai Commissari della statistica agraria calcola che, ai prezzi del 1914, l'incremento annuo rappresenti in media l'1,05 % del valore del 1908.

(5) G. FOTTICCHIA. *Op. cit.* La cifra è riportata dallo ZATTINI nel suo calcolo della produzione agricola italiana.

(6) C. GINI. *Problemi sociologici della guerra*, pag. 253.

Applicando questo rapporto al valore della produzione animale desunto dal calcolo del Franciosa, di cui ci siamo serviti per il calcolo della proprietà terriera, cioè 13.242 milioni di lire, troviamo per il valore del bestiame circa 27 miliardi come dall'altra via prima seguita.

VII.

VALORE DELLA MOBILIA

Questa categoria di beni costituisce uno fra gli elementi più importanti della ricchezza privata, d'altra parte offre difficoltà insormontabili a calcoli diretti.

Ci aiuterà il calcolo congetturale, che, eseguito per diverse vie, se offre risultati concordanti, non è certo meno sicuro di un calcolo diretto.

Si capisce subito come la congettura debba basarsi su di un rapporto fra il valore dei fabbricati e quello della mobilia e come appunto questo rapporto permetta di passare dall'uno all'altro valore.

La difficoltà si incontra però nella scelta del rapporto, che può variare da paese a paese e da epoca a epoca.

Per l'Inghilterra per esempio i rapporti proposti variano in modo sensibilissimo: accanto al Giffen che stima il valore della mobilia uguale alla metà del valore dei fabbricati, troviamo il Chiozza-Money che la stima uguale ad un sesto. (1)

Lo Stamp tenendo conto di una relazione fra valore locativo e mobilia arriva ad un valore di questa che rappresenta un po' meno di un quarto del valore dei fabbricati. (2)

Infine lo stesso criterio segue il Pupin nel suo calcolo della ricchezza della Francia prima della guerra, (3) ma conforta poi il risultato calcolando un valore medio di mobilia per famiglia. (4)

La via del Pupin pare la più sicura, ed è appunto quella che noi seguiremo.

Cominceremo con l'applicare il rapporto risultante dai calcoli del Prof. C. Gini, alla cifra da noi trovata per il 1928, rappresentante il valore dei fabbricati.

(1) Josiah Stamp. *British incomes and property*, pag. 400, 401.
(2) Josiah Stamp. *Wealth and taxable capacity*, pag. 21.
(3) R. Pupin. *La richesse de la France devant la guerre*, pag. 23.
(4) Id. id., pag. 24.

Tale rapporto per il 1914 risulta di 57,5 %, nel 1925 di poco superiore; precisamente del 61,0 %. (1)

L'aumento del valore degli arredamenti delle case di contadini spiega perfettamente l'aumento verificatosi dal 1914 al 1925, e ci spinge per il 1928 ad adottare lo stesso rapporto del 1925.

Il valore della mobilia raggiungerebbe secondo questa via i 48 miliardi.

Se noi d'altra parte supponiamo che il valore della mobilia sia variato come l'indice generale dei prezzi (646,24-491,36) troviamo un valore alquanto inferiore al precedente: 42 miliardi.

Passiamo ora ad un secondo controllo. Per il 1914 il Prof. C. GINI dava come valore medio di mobilia per famiglia 1500 lire. Ai prezzi attuali considerando il coefficiente d'aumento del valore della mobilia uguale a quello dei prezzi il sopra detto valore rappresenterebbe 7400 lire, che, essendo il numero complessivo delle famiglie 8.696.494, darebbe una cifra totale di 62 miliardi da cui detraendo $^1/_3$ per l'usura avremmo 42 miliardi.

Dando però maggior peso alla prima valutazione crediamo di poter fissare il valore della mobilia in 45 miliardi.

VIII

VALORE DEGLI ALTRI MOBILI

OGGETTI DI ORNAMENTO – VEICOLI – FONDI DI COMMERCIO – GIACENZE ED ATTREZZI DEL COMMERCIO E DELL'INDUSTRIA PRIVATA – ATTREZZI AGRICOLI NON APPARTENENTI AI PROPRIETARI DELLA TERRA

Per calcolare questa categoria di ricchezza ci varremo, come al solito, di due vie: l'una basata sull'estensione dei risultati trovati dal Prof. C. GINI, nei suoi calcoli del 1925, e l'altra che consisterà in una valutazione diretta.

Nel 1908 il GINI valutava la categoria in questione, comprendente fondi di commercio, prodotti agrari minerari, macchine ed attrezzi agrari (2) e quelli commerciali ed industriali sempre di proprietà

(1) C. GINI. *Problemi sociologici della guerra*, pag. 255; *A comparison of the wealth* ecc., pag. 6.

(2) Dobbiamo avvertire che gli attrezzi rurali appartenenti ai proprietari della terra sono compresi nella valutazione di questa come i fabbricati rurali.

Naturalmente anche la nostra valutazione per il 1928 ha la stessa caratteristica essendo basata su quelle dell'autore citato.

privata, veicoli, oggetti di ornamento, in 5 miliardi. (1) Egli stesso osservava che tale cifra doveva rimanere al disotto della realtà dato che il PRINCIVALLE valutava nel 1908 a 5 miliardi solo gli attrezzi del commercio e dell'industria, cifre a cui perveniva stimando a 10 volte il reddito, il valore dei mobili, giacenze ed attrezzi. (2)

Nel 1914, lo stesso Prof. C. GINI valutava a 10 miliardi l'intiera categoria degli altri mobili. (3)

Incominceremo dalla valutazione diretta.

OGGETTI DI ORNAMENTO DELLA PERSONA

Seguiremo nella valutazione particolareggiata, che abbiamo intrapreso, quella fatta per la Francia dal PUPIN, e per il Belgio dal BAUDHUIN.

Il primo autore fissa a 1 miliardo il valore dei «meubles corporels» (35 franchi per testa tenendo conto dell'usura; come spesa d'acquisto 280); (4) a 3 miliardi gli oggetti d'oro, (5) a 1 miliardo gli oggetti di altri metalli preziosi e ancora ad un miliardo le pietre preziose. (6)

Cioè attribuisce alla popolazione al di sopra di 18 anni 36 franchi per testa di pietre preziose e ugualmente per gli oggetti di metallo prezioso non d'oro; per gli oggetti d'oro, infine, 108 franchi a testa, essendo la popolazione al disopra di 18 anni circa sui 28 milioni. (7)

Ora, supponendo che vi sia una correlazione fra il reddito per testa e gli oggetti d'ornamento della persona, essendo nel 1914 il reddito per testa Francese sui 945 franchi e quello Italiano sui 530-557, (8) possiamo fissare le cifre per l'Italia, corrispondenti a quelle di cui sopra per la Francia, in 20 lire per testa di biancheria, in 45 per gli oggetti d'oro, in 22, infine, per testa al di sopra dei 18 anni per gli oggetti in metallo prezioso differente dall'oro, e ugualmente per le pietre preziose.

(1) C. GINI. *L'ammontare e la composizione* ecc., pag. 174.
(2) C. GINI. *L'ammontare e la composizione* ecc., pag. 183.
(3) C. GINI. *Problemi sociologici della guerra*, pag. 256.
(4) PUPIN. *Op. cit.*, pag. 24.
(5) Id. id., pag. 26.
(6) Id. Id.
(7) Id. Id.
(8) BANCA COMMERCIALE ITALIANA. *Movimento economico dell'Italia*, 1927, pag. 360.

Si osserva in più che essendo la composizione per età della popolazione Italiana differente da quella Francese, in modo che il numero degli individui delle classi giovanissime è maggiore, relativamente, nella prima che nella seconda, non possiamo accontentarci della riduzione fatta per passare dal valore della biancheria per testa dei cittadini Francesi, al valore di quella dei cittadini Italiani, poichè naturalmente il fatto che esistano molti bimbi diminuisce il valore per testa della biancheria, indipendentemente da altri fattori.

Per tenere conto di questa ultima circostanza fisseremo a 15 lire il valore di cui abbiamo parlato.

Essendo la popolazione nel 1914 calcolata in 35.597.784, avremo per questa voce 354 milioni.

In quanto agli oggetti d'oro, essendo nel 1914 la popolazione al di sopra di 18 anni 21.443.425, a 45 lire per testa troviamo in cifra tonda 1 miliardo. Troviamo un altro miliardo per gli oggetti di metalli preziosi differenti dall'oro e pietre preziose.

Confrontando la cifra ottenuta con le corrispondenti di altre nazioni vediamo che essa è accettabile.

Considerando quella della Francia vediamo, infatti, che i nostri due miliardi e mezzo vanno d'accordo nel quadro della ricchezza totale italiana, con i sei del PUPIN (1) in quello della ricchezza Francese. Così la cifra data dal BAUDHUIN (2) per il Belgio in 700 milioni non contrasta con la nostra.

Possiamo, ora, con relativa sicurezza passare dalla cifra del '14 a quella del '28.

In totale per il 1914 troviamo 2,5 miliardi.

Per la biancheria nel 1928, in moneta attuale, tenendo conto della diminuzione del potere d'acquisto della moneta in base al livello generale dei prezzi, per la popolazione presente valutata in 40.796.000 mila abitanti abbiamo, allora, 3 miliardi.

Per le altre voci, sempre nel 1928, essendo la popolazione al di sopra dei 18 anni 25,566.853, (3) tenuto conto del livello dei prezzi

(1) R. PUPIN. *Op. cit.*, pag. 26. La ricchezza privata francese secondo questo autore ammontava nel 1914 a 285 miliardi e mezzo.

(2) F. BAUDHUIN. *Le capital de la Belgique et le rendement de son industrie avant la guerre*, pag. 39. Secondo i suoi calcoli la ricchezza privata belga ammontava nel 1914 a 45.938 milioni.

(3) La cifra è stata trovata dando alla popolazione nel 1928 la stessa composizione per età che risulta al censimento del 1921.

in confronto all'anteguerra, troviamo 5.650 milioni per l'oro, e altrettanto per le altre due categorie riunite (L. 220 a testa per gli oggetti d'oro, 110 per ciascuna delle altre due categorie). In totale per il 1928 troviamo 14-15 miliardi.

Giacenze, attrezzi, utensili delle industrie e dei commerci privati

Per valutare questa categoria ci serviremo inizialmente del rapporto fissato dal Principalle fra redditi e valore delle giacenze, utensili ecc. dei commercianti ed industriali, controlleremo poi i risultati con dei confronti.

Secondo il Principalle (1) le giacenze e gli attrezzi del commercio e dell'industria ammonterebbero a 10 volte il reddito netto. Ora a pari giacenze il reddito netto dei privati è superiore a quello delle società, d'altra parte vi è l'evasione per cui il reddito netto accertato è inferiore all'effettivo; se ne può dedurre che, contrabilanciandosi le due influenze, moltiplicando per 10 il reddito netto dei privati industriali e commercianti (Categoria *B*, secondo gruppo, per l'imposta di R. M.) la cifra che si ottiene rappresenta il valore delle giacenze e degli attrezzi che cerchiamo.

Secondo gli accertamenti compiuti nel 1923 dal Ministero delle Finanze, in quell'epoca il reddito dei privati commercianti e industriali saliva a 3.214.482.696 lire. (2)

Detraendo da questa somma il reddito derivante dall'industria del credito e degli appalti, che non hanno quasi giacenze, (45.544.683, per le prime e 799.998.123 per le seconde) (3) otteniamo la cifra da moltiplicare per 10: 2.368.939.890.

Il valore della categoria di ricchezza che stiamo valutando ammonterebbe così a 23.689 milioni; cifra che possiamo arrotondare a 24 miliardi, tenendo conto che l'evasione supera la circostanza che, a pari giacenze, il reddito dei privati è superiore a quello delle società per azioni.

(1) C. Gini. *L'ammontare e la composizione* ecc., pag. 183. Il rapporto trovato dal Principalle si basa su ricerche da lui compiute nei bilanci di società.

(2) Queste cifre ci furono gentilmente comunicate dal Dott. S. Orlandi, il quale a sua volta le ha ricevute dal Ministero delle Finanze.

(3) Id. Id. nota precedente.

Ora bisogna porsi la domanda se questa cifra si possa riferire al 1928, oppure occorra tener conto della variazione dei prezzi dal 1923 al 1928. Dato che l'accertamento del reddito ha un carattere d'invariabilità per un certo periodo di anni, parrebbe lecito di non tener conto delle variazioni dei prezzi, in quanto, essendo il reddito accertato il reddito ritraibile in media, si può supporre che anche il valore delle giacenze corrispondenti abbiano questo carattere di media.

Però, dato l'aspetto radicale della mutazione apportata ai valori dalle variazioni dei prezzi, riteniamo prudente di tener conto del fatto che il livello dei prezzi era, appunto nel 1923, superiore a quello del 1928, essendo precisamente l'indice del livello dei prezzi nel 1923 530.78 e quello del 1928 491,36. (1)

Con una semplice proporzione arriviamo alla cifra di 22 miliardi, che giudichiamo attendibile.

In essa, dobbiamo osservare, sono compresi anche i valori degli attrezzi agricoli non appartenenti ai proprietari e quindi tassabili nel loro reddito sotto l'imposta di R. M.

Passiamo ora ad un controllo:

Il numero degli articoli di ruolo, cioè quello delle singole aziende, saliva all'epoca dell'accertamento a 794.727, da cui detraendo il numero di quelle aziende che non abbiamo considerato (5284 per le aziende di credito, 16.400 per gli appalti), abbiamo 773.043 aziende. Il valore medio verrebbe ad essere secondo i nostri calcoli 27.680 lire.

Il Pupin, per la Francia nel 1914 valutava questa categoria di beni a 18 miliardi, cioè 20.600 franchi in media per azienda. (2)

Ora si capisce come questa cifra, data la molto maggiore ricchezza della Francia e la svalutazione della moneta, si accordi con quella da noi trovata, e, anzi, faccia sospettare che questa ultima sia troppo bassa.

(1) Indici del Consiglio Provinciale dell'Economia di Milano.

(2) R. Pupin. *Op. cit.*, pag. 19, 20, 21, 22.

IX.

VEICOLI DI TERRA E DI MARE APPARTENENTI A PRIVATI

Un esauriente materiale ci è dato dalle iscrizioni nel pubblico registro automobilistico al 31 dicembre 1928. (1)

In base a tale iscrizione all'epoca detta si avevano in Italia:

Autovetture private e pubbliche	143.000
Autoveicoli propriamente detti	56.000
Motocicli e motocarrozzette	64.403

Per gli autoscafi prendiamo la cifra data per il 1927 dalla Direzione generale del Demanio e delle Tasse, cioè 1584. (2)

Non calcoleremo affatto fra la ricchezza privata il valore degli autoveicoli propriamente detti che sono costituiti principalmente da autocarri, trattrici, ecc. cioè veicoli che generalmente appartengono a società per azioni. D'altra parte quei veicoli di questa categoria che non appartengono a tale categoria, verranno compensati da quelli dell'altra appartenenti a società.

Si tratta ora di dare un prezzo venale medio ad ogni categoria. Lo faremo in base al parere di competenti. Naturalmente il risultato sarà largamente approssimativo, ma data l'ampiezza del campo d'oscillazione permesso alla cifra esprimente la ricchezza totale, tale approssimazione molto larga non nuocerà affatto all'insieme del calcolo.

Per le autovetture il prezzo medio venale, tenuto conto del deprezzamento correlativo alle varie cilindrate e alla diffusione di queste, pare aggirarsi sulle 16.000 lire; per i motocicli e motocarrozzette sulle 5.000 lire, per i motoscafi sulle 20.000.

Facendo il calcolo risulta:

Autovetture	L.	2.336.000.000
Autoscafi	»	31.680.000
Motocicli ecc.	»	322.015.000
TOTALE . . .	L.	2.889.695.000

(1) « Bollettino mensile di Statistica ». Aprile 1929, pag. 377.

(2) « Annuario Statistico Italiano », 1928, pag. 239.

Tenendo conto che i prezzi medi fissati sono probabilmente più bassi dei reali, che non sono stati calcolati i valori delle biciclette a motore, dei canotti a remi e a vela, delle barche da pesca, delle carrozze ecc. non crediamo esagerato di portare il totale a 3 miliardi. (1)

Tirando le somme delle varie componenti la categoria « altri mobili » abbiamo un totale di 41-42 miliardi.

Questa cifra, tenendo conto della rivalutazione della moneta dal 1925 ad ora, e della maggiore ricchezza nata in questi anni, si accorda perfettamente con quella data dal Prof. Gini, appunto per il 1925, in 45 miliardi. (2)

X.

VALORE DEI DEPOSITI

Secondo la situazione delle Casse Postali di risparmio, il credito dei depositanti, alla fine del dicembre 1928 (3) presso le casse stesse ammontava a 10.486 milioni.

Nelle Casse di risparmio ordinarie, il credito dei depositanti per somme collocate a risparmio, in conto corrente, in buoni fruttiferi e depositi di altra specie ammontava alla stessa epoca a 15.626,2 milioni. (4)

Infine presso gli Istituti principali di credito, nell'ottobre 1928, il credito dei depositanti ammontava a 3.711,5; e nella stessa epoca si avevano nelle banche popolari 984,2 e presso le banche regionali 4.027,2 milioni. (5)

A questi dobbiamo aggiungere infine il credito dei correntisti dei conti correnti postali che saliva alla fine del 1928 a 218,8 milioni. (6)

Sommando le varie cifre esposte abbiamo 35.000,0 milioni in cifra tonda.

(1) Da questo valore naturalmente sono esclusi i valori dei veicoli dell'esercito, marina, ministeri, ambasciate estere.

(2) C. Gini. *A comparison of the wealth* ecc., pag. 6.

(3) « Bollettino mensile di Statistica », marzo 1929, pag. 270.

(4) Id. Id.

(5) Id. Id.

(6) C. Gini. *Problemi sociologici della guerra*, pag. 254.

Di questa somma i nove decimi si possono ritenere appartenenti a privati, come ha provato il Prof. C. GINI, (1) quindi possiamo valutare la categoria di ricchezza privata che stiamo trattando in cifra tonda a 31,5 miliardi.

XI.

VALORE DEL DENARO

Al dicembre 1928 la circolazione complessiva (bancaria e di stato) ammontava a 17.456,4 milioni. (2)

Le monete metalliche ammontavano al 30 novembre 1928 a 1.733.570.450. Naturalmente non tutta questa massa di circolante va considerata all'attivo della fortuna privata, poichè dobbiamo detrarre le giacenze presso il Tesoro, la Cassa Depositi e Prestiti, gli Istituti di emissione, presso le società industriali e commerciali, nonchè presso le casse degli Uffici ministeriali, enti pubblici ecc.

La detrazione presenta delle forti difficoltà, ma d'altra parte sarà possibile girare l'ostacolo.

In realtà noi possiamo supporre che la massa di moneta in mano dei privati siasi andata contraendo, nella stessa misura della circolazione totale dal 1925 al 1928.

In altre parole possiamo supporre che la distribuzione fra privati e non privati della moneta sia rimasta inalterata dal 1925 al 1928.

Se si trattasse unicamente della moneta formante la circolazione bancaria, il calcolo si potrebbe eseguire riducendo la massa di moneta in mano dei privati nel 1925 secondo gli indici costruiti dall'Istituto Centrale di Statistica, ma la circostanza dell'emissione di moneta metallica impedisce di eseguirlo in siffatto modo.

Nel 1925 la circolazione totale comprendeva 849.889.424 di moneta metallica, più 21.449,6 milioni di circolazione cartacea complessiva. (3)

In totale 22.299,5 milioni.

(1) C. GINI. *Problemi sociologici della guerra*, pag. 273.
(2) « Bollettino mensile di Statistica », gennaio 1929.
(3) « Annuario Statistico Italiano », 1922-25, pag. 320-321.

Di questi il Prof. C. Gini calcolava in mano dei privati soltanto 7 miliardi, cioè il 31.39 %. (1)

Applicando questa percentuale alle cifre del 1928 troviamo che il denaro in mano dei privati in questo anno raggiungeva i 6 miliardi.

XII.

VALORE DEI BENI DI RICCHEZZA MOBILE PER DESIGNAZIONE DELLA LEGGE

Titoli di debito pubblico

Dobbiamo, conformemente al criterio stabilito, calcolare all'attivo della ricchezza privata i titoli di debito pubblico, ma naturalmente soltanto quelli situati nelle mani di privati altrimenti andremmo incontro a ripetizioni, poichè parte di detti titoli rappresentano investimenti di capitale azionario di società per azioni, oppure dei depositi presso istituti di risparmio, o infine di fondi appartenenti a enti di diritto pubblico.

Per determinare la parte del debito pubblico piazzata fra privati ci varremo di una proporzione, in mancanza di altra più recente, fissata dal Principalle per il 1903 e usata dal Prof. C. Gini anche per il 1914. (2) Secondo tale proporzione il 55 % dei titoli di debito pubblico si trovano nelle mani di privati.

Per il 1928, essendo, alle quotazioni di borsa, il debito pubblico in totale ammontante a 70.261 milioni, applicando la percentuale del 55 %, abbiamo 38.644 milioni.

(1) C. Gini. *A comparison of the wealth* ecc., pag. 6.

Lo stesso A., per il 1914 trovava che la proporzione in discorso era del 36,84 %.

Da allora sarebbe, quindi, diminuita la frazione della massa di circolante situata in mano dei privati. Le ragioni di questo fenomeno si trovano, probabilmente nella diminuzione del tesoreggiamento, nella maggior tensione della circolazione procurata dalla svalutazione della moneta.

Ad ogni modo, sembra opportuno adottare per il 1928 la proporzione del 1925, piuttosto che quella del 1914, poichè le condizioni dell'anno scorso sono molto più prossime a quelle del 1925 che a quelle del '14.

(2) C. Gini. *Problemi sociologici della guerra*, pag. 254.

Azioni, certificati, cartelle, obbligazioni

Il valore di tali titoli escluse le obbligazioni ammontava come valore nominale al 31 ottobre 1928 a 44.488.642.445 lire. (1)

Questa cifra è naturalmente ben lontana dal rappresentare la ricchezza italiana in azioni, poichè per raggiungere questa è necessario tener conto delle quotazioni che hanno in Borsa i titoli quotati, e comunque il valore effettivo, differente da quello nominale dei titoli in genere.

Ci sarà d'aiuto per il passaggio dalla cifra sopra nominata a l'effettiva l'indice calcolato dal Guarnieri che ci dà la quotazione generale media per ogni 100 lire versate.

Tale indice è per il 1928 in media 181,41.

Variando in proporzione la cifra che dà il capitale nominale otteniamo 80.706.000.000.

Le obbligazioni alla fine del 1927, comprendendovi quelle ferroviarie, raggiungevano la cifra di 2.409.286.357 come valore nominale (2) a cui si devono aggiungere ancora 713 milioni (3) di nuove emissioni per il 1928 cioè in totale 3.122.286.357. Per passare da questa cifra al valore effettivo abbiamo calcolato una quotazione media ogni 100 lire nominali, basandoci sulla quotazione di alcuni tipi di obbligazioni, e mediante una media ponderata, in cui i pesi sono costituiti dall'ammontare delle obbligazioni emesse dalle società corrispondenti.

Abbiamo esteso poi questa media a l'ammontare totale delle obbligazioni.

Tale quotazione media ogni 100 lire nominali risulta di 65 e in conseguenza il valore delle obbligazioni è di 2 miliardi in cifra tonda. Concludendo la ricchezza rappresentata da titoli azioni, obbligazioni, sarebbe di 82,7 miliardi.

Dobbiamo ora valutare le azioni Italiane collocate all'Estero. Sulla base dei dati forniti dalla Banca Commerciale Italiana (4) possiamo stimare l'ammontare di tali azioni, a una quotazione media dell'annata 1928, sul mezzo miliardo. Questa cifra è inferiore alla realtà, sicuramente, poichè molte azioni sono sfuggite al calcolo presentato dalla banca sopranominata ma possiamo supporre che l'eccedenza sulla cifra da noi assunta sia compensata dall'ammontare di azioni

(1) Associazione fra le Società Italiane per azioni. *Notizie statistiche sulle società per azioni*, 1929.

(2) Id. Id.

(3) Banca Commerciale Italiana. *Movimento economico Italiano*, 1928, pag. 91.

(4) Id. Id.

estere collocate in Italia. In modo più esatto possiamo stabilire l'ammontare delle obbligazioni nostre collocate all'estero, che alle quotazioni delle borse estere su cui sono trattate, rappresentano una cifra di 2.750 milioni di lire.

La ricchezza totale in azioni ed obbligazioni risulta così di 79,5 miliardi.

Di questi si può ammettere che i due terzi siano in mano di privati, vale a dire si possono assegnare ad essi 52 miliardi di titoli, azioni, ecc. (1)

Detta cifra si accorda con quella dal 1925 se si tien conto del forte incremento del capitale nominale azionario, di lire 8 miliardi, (2) che giustifica la differenza tra l'attuale cifra e i 40 miliardi calcolati per il 1925 dal Prof. Gini (3).

Prestiti nazionali esteri collocati in Italia

Secondo i dati della Banca Commerciale Italiana, il valore dei titoli dei prestiti esteri di proprietà italiana salirebbe a circa 1 miliardo. (4)

Supponendo che di questo soltanto il 55 % sia nelle mani dei privati, come accade per i prestiti nazionali, dobbiamo aggiungere alla ricchezza privata Italiana ancora mezzo miliardo.

XIII.

VALORE DELLE PASSIVITÀ

Dobbiamo infine calcolare le passività della ricchezza privata.

Principale fra queste l'eccedenza del debito ipotecario sul credito ipotecario.

Molti autori, come il Baudhuin, il Pupin, ammettono che tale passività non esista in quanto suppongono che il debito ipotecario costituisca un credito di ugual valore per altri privati.

Ciò non risponde a realtà, perchè creditori e debitori non è detto che siano sempre privati, ma può essere che il credito e il debito siano

(1) C. Gini. *Problemi sociologici della guerra*, pag. 321.
(2) Associazione fra le Società per azioni. *Op. cit.*
(3) C. Gini. *A comparison of the wealth* ecc., pag. 6.
(4) Banca Commerciale Italiana. *Op. cit.*

in mano ora di privati ora di persone di diritto pubblico. Effettivamente, dalla statistica del debito ipotecario, eseguita nel 1910, per l'Italia, risulta che il valore dei debiti ipotecari a carico di privati supera di molto quello dei crediti ipotecari a favore di privati.

È quindi necessario calcolare sia l'attivo sia il passivo di tale debito.

La mancanza di dati recenti toglie però la possibilità di condurre il calcolo con esattezza e sicurezza. Congetturalmente potremo riuscire soltanto a fissare l'ordine di grandezza di tale passività.

Venendo al calcolo: possiamo supporre che il rapporto del credito ipotecario, e debito ipotecario al valore della proprietà fondiaria sia rimasto immutato dal 1914.

Ciò equivale ad ammettere che il debito e il credito ipotecario siano cresciuti come il valore dei terreni e dei fabbricati urbani, sia per opera dell'aumento dei prezzi, sia per opera dell'aumento della quantità di tali beni.

È evidente che tale ipotesi è molto grossolana, ma d'altra parte è l'unica a nostra disposizione.

Allora partendo dai dati per il 1910, che il Prof. C. Gini giudica valevoli anche per il 1914, (1) possiamo effettuare senz'altro il calcolo.

Nel 1910 la statistica del debito ipotecario, trovava che il credito ipotecario dei privati ammontava a 2.325.113.906; e il debito ipotecario a 3.836.513.696 lire. (2)

Rapportando tali valori al valore globale dei fabbricati e dei terreni, secondo il calcolo del Prof. C. Gini, otteniamo un rapporto di 3,63 % per il primo, e del 5,99 % per il secondo. Applicando queste percentuali ai valori per il 1928, si ottiene che il credito ipotecario salirebbe a 8.700 milioni e il debito a 14.500 milioni in cifra tonda.

La passività ammonterebbe quindi a 6 miliardi.

Conformemente ai criteri seguiti, nei suoi calcoli, dal Prof. C. Gini, di altre passività a carico dei privati non valuteremo poi che il debito cambiario, supponendo che altre specie di debiti vengano compensate da crediti non ipotecari. (3)

Anche per il debito cambiario dobbiamo ricorrere a una ipotesi di cui l'attendibilità è molto incerta, ma che, d'altra parte, è l'unica a nostra disposizione.

(1) C. Gini. *Problemi sociologici della guerra*, pag. 256.

(2) « Annuario Statistico Italiano », 1914, pag. 365.

(3) C. Gini. *Problemi sociologici della guerra*, pag. 338. *L'ammontare e la composizione* ecc., pag. 179.

Possiamo supporre che il rapporto tra valore del debito cambiario e l'ammontare dei depositi sia nel 1928 quello stesso che si verificava nel 1914. Tale rapporto per il 1914 era di 21% (1) ed applicato alla cifra attuale dei depositi darebbe 6 miliardi circa. In totale le passività della ricchezza privata ammonterebbero a 12 miliardi. D'altra parte se si considera che la guerra ha creato una congiuntura favorevole al riscatto delle ipoteche si capisce come la nostra ipotesi sia eccessivamente pessimistica, e quindi la cifra trovata, troppo alta. Per queste ragioni crediamo giusto ridurla a 10 miliardi incontrandoci così con la valutazione del Prof. C. Gini per il 1925. (2)

Effettivamente, tenuto conto della rivalutazione della lira da una parte, dell'aumentato numero delle ipoteche per la ripresa dell'agricoltura dall'altra, non pare assurdo questo risultato.

XIV.

CONCLUSIONI

Riepilogando i risultati dei calcoli eseguiti possiamo costruire la seguente tabella:

La ricchezza privata degli italiani nel 1928

(*In milioni*)

Terreni	155.000
Fabbricati urbani	80.000
Cave-miniere	5.000
Mobilia	45.000
Altri mobili	45.000
Bestiame	27.000
Titoli di debito pubblico	38.000
Titoli di debiti esteri	500
Azioni, obbligazioni	52.000
Depositi	31.500
Danaro	6.000
Totale lordo	485.000
Passività	10.000
Totale netto	475.000

(1) C. Gini. *Problemi sociologici della guerra*, pag. 254.

(2) C. Gini. *A comparison of the wealth* ecc., pag. 6.

Concluso così il nostro calcolo viene fatto di domandarsi se la ricchezza privata degli italiani è cresciuta o no dall'anteguerra.

L'interrogazione è legittima e merita una risposta.

Dobbiamo ricordare innanzi tutto che per compiere il paragone che c'interessa è necessario esprimere la ricchezza presente in termini della moneta prebellica, dividere cioè per il livello generale dei prezzi il totale del nostro calcolo.

Eseguendo la semplice operazione ora indicata troviamo che in lire d'anteguerra la ricchezza privata degli italiani è di 97 miliardi, cioè nettamente inferiore a quella del 1914.

Il risultato raggiunto provoca meraviglia, poichè ognuno ha la chiara impressione che la ricchezza italiana in beni non è diminuita certamente, ma piuttosto cresciuta dall'anteguerra. Questa impressione effettivamente corrisponde a realtà, come gli elementi quantitativi dei calcoli precedenti possono sufficientemente dimostrare.

Possiamo quindi concludere che l'espressione monetaria della fortuna privata degli italiani è diminuita dall'anteguerra ad ora ma che la quantità di beni è piuttosto cresciuta. La spiegazione di questo fenomeno, come mostrano le lunghe analisi che abbiamo fatto della variazione dei prezzi della terra, dei fabbricati e di altri beni, è in quelle circostanze che il Prof. C. Gini descrive, in questi termini: (1)

« Quando si parla di ricchezza da valutare, si intende infatti riferirsi alla ricchezza mobiliare o immobiliare: ora è chiaro che un bene immobile o mobile, pure dando con lo stesso lavoro lo stesso prodotto, avrà sul mercato un prezzo maggiore o minore a seconda che di tale prodotto è maggiore o minore la frazione che viene attribuita al proprietario...

...Variazioni siffatte non alterano, se non indirettamente, la potenzialità economica o la massa o il flusso dei beni della nazione, ma possono invece alterare fortemente i risultati della valutazione monetaria della ricchezza ».

Questa osservazione ci permette di andare più a fondo nelle nostre conclusioni. Le variazioni nella distribuzione dei redditi, a cui accenna il Prof. C. Gini, se da una parte diminuiscono il valore dei beni, evidentemente, dall'altra, innalzano la retribuzione del lavoro.

Ciò significa che la ricchezza personale è aumentata.

(1) C. Gini. *La ricchezza comparata delle nazioni.* « Nuova Antologia », 16 aprile 1926.

Possiamo quindi finire il nostro calcolo dicendo che in Italia dall'anteguerra ad ora è diminuita la ricchezza materiale espressa in termini monetari, ma è aumentata la ricchezza personale, che d'altra parte è un elemento essenziale della potenzialità economica della nazione. È per questo che, intendendo la parola ricchezza nel suo significato più comprensivo, possiamo affermare che la ricchezza dell'Italia è, nel suo complesso, cresciuta.

ROMA, 1929 - VII — STABILIMENTO TIPOGRAFICO DITTA CARLO COLOMBO.